



UNIVERSIDADE DA CORUÑA

Máster en Biología Molecular, Celular y Genética

Memoria del Trabajo de Fin de Máster

**Análisis del ADN repetitivo del genoma del lacértido
Iberolacerta monticola mediante *Next Generation
Sequencing***

**Análise do ADN repetitivo do xenoma do lacértido
Iberolacerta monticola mediante *Next Generation
Sequencing***

**Analysis of repetitive DNA of the lacertid *Iberolacerta
monticola* genome through *Next Generation Sequencing***

López Díaz, Iñaki

Curso 2020/2021. 18 de junio, 2021

Firmado por LOPEZ DIAZ,
IÑAKI (FIRMA) el día
18/06/2021 con un
certificado emitido por AC
DNIE 006

Director Académico: Naveira Fachal, Horacio

1 NAVEIRA FACHAL
HORACIO - DNI
32423371H



Firmado digitalmente por NAVEIRA
FACHAL HORACIO - DNI 32423371H
Fecha: 2021.06.18 21:11:09 +02'00'

Índice

Resumen	3
Resumo	3
Summary	3
1. Introducción.....	4
2. Objetivos	11
3. Materiales y métodos	11
4. Resultados	15
5. Discusión	34
6. Conclusiones	36
7. Bibliografía	37

Resumen

Iberolacerta monticola es una de las especies pertenecientes a la familia Lacertidae, uno de los grupos de reptiles más diversos del Paleártico Occidental. Su composición de ADN satélite genómico ha sido estudiada con anterioridad mediante técnicas clásicas de análisis de estos elementos repetitivos, basadas en la restricción parcial del ADN genómico. El objetivo de este trabajo es estudiar con *NGS* y la herramienta *RepeatExplorer2* la exactitud de los resultados obtenidos con los métodos clásicos. Los resultados demuestran la presencia de las familias de ADN satélite *HindIII* y *TaqI*, y además confirman la existencia de diferencias en las proporciones de las subfamilias I y II de *HindIII* en *I. monticola*. También se ha comprobado que este procedimiento permite la obtención de otras secuencias repetitivas de interés.

Resumo

Iberolacerta monticola é unha das especies pertencentes á familia Lacertidae, un dos grupos de réptiles máis diversos do Paleártico Occidental. A súa composición de ADN satélite xenómico foi estudada con anterioridade mediante técnicas clásicas de análise destes elementos repetitivos, basadas na restrición parcial do ADN xenómico. O obxectivo deste traballo é estudar con *NGS* e a ferramenta *RepeatExplorer2* a exactitude dos resultados obtidos cos métodos clásicos. Os resultados demostran a presenza das familias de ADN satélite *HindIII* y *TaqI*, e ademais confirman a existencia de diferencias nas proporcións das subfamilias I e II de *HindIII* en *I. monticola*. Tamén comprobouse que este procedemento permite a obtención de outras secuencias repetitivas de interese.

Summary

Iberolacerta monticola is one of the species belonging to the Lacertidae family, one of the most diverse groups of reptiles in the Western Palearctic. Its genomic satellite DNA composition has been previously studied using classical analysis techniques for these repetitive elements, based on partial restriction of genomic DNA. The objective of this work is to study with *NGS* and the *RepeatExplorer2* tool that the results obtained with classical methods are valid. The results demonstrate the presence of the satellite DNA families *HindIII* and *TaqI* and confirm the existence of differences in the proportions of subfamilies I and II of *HindIII* in *I. monticola*. It has also been found that this procedure allows the obtaining of other repetitive sequences of interest.

1. Introducción

La familia Lacertidae es una de las familias de reptiles más diversas y expandidas del Paleártico Occidental (García-Porta *et al.*, 2019). Esta familia posee unas 340 especies, entre las que están las del género *Iberolacerta*, que agrupa a *Iberolacerta monticola*, *I. cyreni*, *I. galani* e *I. martinezricai*, de las regiones centro-occidentales de la Península Ibérica, *I. aranica*, *I. aurelioli* e *I. bonnali*, exclusivas de los Pirineos. También pertenece a este grupo *I. horvathi*, presente en los Alpes Orientales y Dináricos (Giovannotti *et al.*, 2014).

Iberolacerta monticola es una especie nativa de la península Ibérica perteneciente a la familia Lacertidae. Sus dos subespecies habitan en regiones de elevada altitud: *I. monticola monticola*, exclusiva de la Serra da Estrela, en Portugal, e *I. monticola cantabrica*, presente en la Cordillera Cantábrica y Galicia (Galán *et al.*, 2007; Remón *et al.*, 2013). El grado de aislamiento de esta especie a lo largo de su área de distribución es muy variable: existen regiones donde el aislamiento es mínimo y sus poblaciones son grandes e hibridan (e. g., Cordillera Cantábrica), mientras que en otras, sus poblaciones presentan pocos miembros y viven muy aisladas unas de otras debido a lo fragmentado que está su hábitat (Galán *et al.*, 2007; Remón *et al.*, 2013). Esta fragmentación es debida a la destrucción de hábitat a causa de las actividades humanas y conlleva grandes peligros para estas pequeñas poblaciones (Remón *et al.*, 2013). Estas poblaciones de reducido tamaño se ven muy afectadas por procesos estocásticos ambientales (deriva génica) y endogamia, los cuales pueden llevarlas a la extinción con gran facilidad (Remón *et al.*, 2013).

En este trabajo, hemos realizado un análisis de las secuencias de ADN repetitivo presentes en su genoma utilizando las modernas técnicas de *Next Generation Sequencing* (NGS).

Los genomas eucariotas están formados en gran medida por ADN no codificante (Choi *et al.*, 2020). Este tipo de ADN puede abarcar grandes regiones, caracterizadas por un elevado grado de variación entre individuos (Miller & Therman, 2001). En cuanto a su ubicuidad en el árbol de la vida, su abundancia en los genomas de los organismos vivos varía enormemente entre especies, llevando a la aparición de fenómenos como la paradoja del valor C (Kubis *et al.*, 1998). Se ha averiguado que la solución a la paradoja del valor C era la cantidad variable de ADN no codificante (“junk DNA”) presente en los genomas de los organismos vivos. Aunque en un principio se denominó como ADN “basura”, hoy en día no se descarta que pueda jugar algún importante papel en la organización del genoma y la regulación de la expresión génica

(Choi *et al.*, 2020). Uno de los tipos de ADN no codificante que aparecen en los genomas de todos los organismos eucariotas es el ADN repetitivo (Miller & Therman, 2001).

El ADN repetitivo se puede dividir en 2 grupos en función de su organización genómica y localización cromosómica. Estos 2 grupos son los ADNs repetidos en tandem, formados por ADN satélites, minisatélites y microsatélites, y los ADNs repetidos dispersos, formados por los elementos transponibles (Kubis *et al.*, 1998). Los elementos transponibles son secuencias muy importantes que juegan un papel fundamental en la evolución de la estructura y función de los genomas, pudiendo aportar tanto ventajas adaptativas a sus hospedadores por creación de novedades genéticas como desventajas causadas por anomalías genéticas causadas por la presencia de estas secuencias (Boissinot *et al.*, 2019).

Los ADN satélites son secuencias simples formadas por unidades de repetición de entre 2 a 2000 pares de bases de longitud (Miller & Therman, 2001). En función del tamaño de su unidad de repetición, los ADN satélites se suelen clasificar en: macrosatélites (unidades de repetición del orden de kilobases), satélites propiamente dichos (de tamaño y complejidad muy variable, generalmente formados por motivos repetitivos más pequeños, presentes generalmente en centrómeros), minisatélites (unidades de repetición de unos 15 pares de bases, en regiones subteloméricas) y microsatélites (unidades de repetición de entre 2 y 6 pares de bases, distribuidos aleatoriamente por el genoma; Thakur *et al.*, 2021).

Además de esta clasificación en función del tamaño de repetición, los ADN satélites funcionales también pueden dividirse en: satélites centroméricos, pericentroméricos y teloméricos y subteloméricos (Thakur *et al.*, 2021).

Aunque existen especies (*S. cerevisiae*) en las que los centrómeros presentan una secuencia no repetitiva, la mayoría de eucariotas poseen secuencias de ADN satélites centroméricos (Thakur *et al.*, 2021). En cuanto a los ADN satélites pericentroméricos, este tipo de secuencia repetitiva es el tipo de ADN satélite más común en los genomas eucariotas (Thakur *et al.*, 2021). Muchos eucariotas reúnen este tipo de ADN en estructuras llamadas cromocentros durante la interfase (Thakur *et al.*, 2021). Por último cabría hablar sobre los ADN satélites teloméricos y subteloméricos. Estos elementos repetitivos se suelen clasificar en microsatélites (telómeros) y satélites (regiones subteloméricas; Thakur *et al.*, 2021). Los ADN satélites teloméricos suelen ser mantenidos por la telomerasa, pero existen grupos como *Drosophila* en los que no se utiliza esta enzima, sino que se emplea a un elemento retroviral domesticado (Thakur *et al.*, 2021).

Un ejemplo de ADN satélite es el ADN α -satélite de primates, con diferencia la familia de ADN satélite más estudiada. Este satélite está formado por largos monómeros de 170 pares de bases repetidos varias miles de veces (Miller & Therman, 2001). Feliciello *et al.* (2020) han estudiado a este tipo de ADN repetitivo y han observado que alrededor del 80% de sus secuencias se localizan en las regiones centroméricas y pericentroméricas de los cromosomas formando clústers de repeticiones separadas por menos de 1 kb entre ellas. El resto de α -satélites son secuencias repetitivas individuales localizadas de manera dispersa por las regiones eucromáticas (Feliciello *et al.*, 2020). Este ADN satélite es también un componente de los centrómeros humanos y está presente en todos sus cromosomas (Miller & Therman, 2001).

En cuanto al descubrimiento de los ADN satélites, estos fueron detectados por primera vez en los años 60 (Kit, 1961). Durante esta época se inventó la ultracentrifugación en gradiente de densidad, una técnica que probó ser muy útil en la caracterización del ADN (Meselson *et al.*, 1957). Meselson *et al.* (1957) utilizó esta técnica con ADN obtenido del timo de un ternero, observando una distribución asimétrica del material genético a lo largo de un gradiente de CsCl. Este fenómeno ponía en relevancia la heterogeneidad del material genético (Figura 1; Meselson *et al.*, 1957).

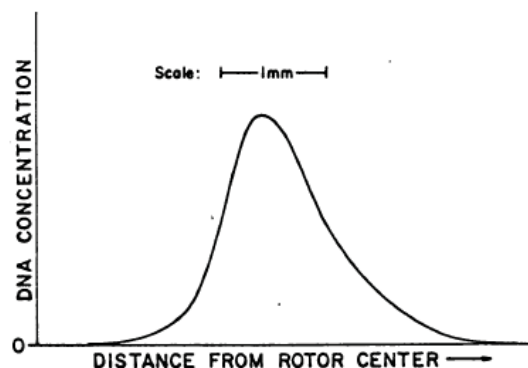


Figura 1: imagen tomada de Meselson *et al.* (1957). Se representa la concentración de ADN (en unidades relativas de concentración) frente a la distancia al rotor de la centrifugación (en mm).

Cinco años después, Schildkraut *et al.* (1962) confirman que el porcentaje de G+C presente en la secuencia condiciona la densidad del ADN. Además, observa con mayor claridad la presencia de ADN satélite en muestras de timo de ternero y espermatozoides de salmón mediante análisis por absorción (Figura 2; Schildkraut *et al.*, 1962).

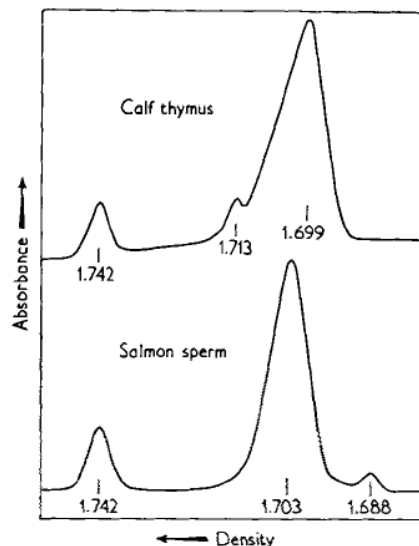


Figura 2: imagen tomada de Schildkraut *et al.* (1962). En ella se muestra una gráfica en la que se enfrenta absorbancia frente a densidad del gradiente de CsCl.

Kit (1961) fue el primer investigador en acuñar el término “ADN satélite” para describir al ADN presente en esas bandas. Él analizó el material genético de varios animales en busca de estas bandas, detectándolas por ejemplo en ratón con una densidad de 1.690 g/cm^3 , frente a 1.701 g/cm^3 del ADN genómico (Kit, 1961).

Esta técnica pronto quedó obsoleta, ya que por ejemplo, no era capaz de discernir diferentes tipos de ADN repetitivo, pasándose a otros métodos con una mayor capacidad de análisis.

Uno de estos nuevos métodos fue el estudio de la cinética de renaturalización mediante el uso de curvas C_{ot} . Esta técnica se basa en que el ADN monocatenario es capaz de absorber una cantidad mayor de luz UV que el ADN bicatenario. Esta diferencia permite cuantificar *in situ* la cantidad de ADN monocatenario presente en un entorno en el que se esté renaturalizando ADN desnaturalizado (Pierce, 2002). Estos valores pueden mostrarse en una gráfica que muestre la proporción de ADN monocatenario frente a concentración por tiempo (ya que la tasa de reasociación del ADN depende de su concentración y del tiempo; Pierce, 2002). Gracias a esta técnica se confirmó que los genomas eucariotas están formados por secuencias que se pueden clasificar en 3 tipos diferentes de clases cinéticas (Hake & Walbot, 1980). Las clases cinéticas son: de reasociación muy rápida, de reasociación rápida y de reasociación lenta (Figura 3; Hake & Walbot (1980) las encontraron en *Zea mays*).

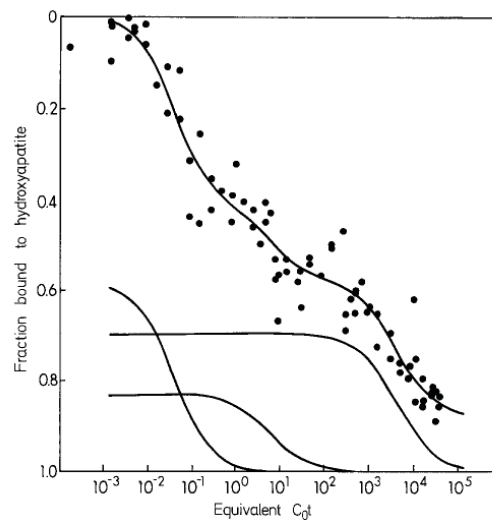


Figura 3: figura extraída de Hake & Walbot (1980). En ella se muestra la curva C_0t obtenida con fragmentos genómicos de *Zea mays* (ADN unido a hidroxiapatita frente a C_0t equivalente). Se puede observar los 3 componentes que forman a la curva principal (curvas sin puntos y curva con puntos, respectivamente).

Debido a que el ADN repetitivo está presente en un mayor número de copias que el de copia única, este es el primero en reasociarse (Britten & Kohne, 1968). Por ello, primero se reasocia el ADN altamente repetitivo, después el medianamente repetitivo y por último, el de copia única (Pierce, 2002).

Uno de los misterios a resolver en cuanto a este tipo nuevo de ADN fue cómo se replica y expande por el genoma. Se formularon varias hipótesis para explicarlo. La primera de ellas es la de replicación saltatoria, realizada por Britten y Kohne (1966, 1968). Según su hipótesis, la secuencia satélite primero sufriría un proceso de duplicación (por *slippage* de la ADN polimerasa o por amplificación por círculo rodante), tras el cual las copias resultantes se integran por el genoma (Beridze, 2013). Si alguna de estas copias se asocia por azar a un elemento genético favorable, esta podrá así fijarse en la población con facilidad (Beridze, 2013).

Otra hipótesis que se planteó posteriormente fue la de por entrecruzamiento desigual, propuesta por Smith (1976). Smith (1976) plantea que las secuencias repetitivas (como los ADN satélites) surgen por entrecruzamiento desigual aleatorio entre cromátidas hermanas. Este tipo de entrecruzamiento produce duplicaciones (en *tandem*) en una de las hebras acompañadas de deleciones en la otra. Si se realizan varias rondas de entrecruzamiento desigual, podrían producirse grandes regiones duplicadas con relativa facilidad. Sin embargo, todos estos cambios pueden ser eliminados rápidamente por la selección natural si afectan a secuencias

codificantes debido a que producen graves alteraciones en la secuencia original (Smith, 1976). Aunque la secuencia original no esté sometida a selección natural, sí que puede verse afectada por procesos estocásticos de deriva genética, pudiendo ser eliminada o pudiendo fijarse en la población por azar (Smith, 1976).

Por último está la hipótesis del mecanismo de formación de 2 pasos, escrita por Walker (1971). Esta hipótesis parte de que la formación de la mayoría de ADN satélites simples se forman en un proceso dividido en 2 partes: primero, se forma un fragmento de aproximadamente 200 pares de bases y tras esto, esta secuencia se duplicaría un elevado número de veces mediante amplificación (Beridze, 2013). En cuanto a los mecanismos por los que se llevarían a cabo estos dos procesos, Southern (1975) propuso que el primer paso se efectuaría a través de replicación saltatoria y el segundo, mediante entrecruzamiento desigual.

A finales de los 70, los métodos de estudio del ADN satélite por centrifugación en gradiente de CsCl y por análisis de curvas *C_{ot}* quedaron obsoletos, siendo sustituidos por el análisis de fragmentos de ADN genómico obtenidos por la digestión parcial con enzimas de restricción (Garrido-Ramos, 2017). Esta nueva metodología era mucho más sencilla que sus predecesoras y permitió el estudio de nuevos ADN satélites y la obtención de familias de ADN satélite. Esta técnica se realiza de la siguiente manera: primero, se digiere parcialmente el ADN genómico total purificado con una enzima de restricción; hecho esto, se realiza una electroforesis en gel de agarosa con la intención de separar por tamaño a los fragmentos de ADN obtenidos tras la digestión. Si después se tiñe el gel con algún colorante inespecífico (e. g., bromuro de etidio), se puede observar la aparición de una banda bien definida en medio de un largo *smear* (Figura 4; Garrido-Ramos, 2017). Esta banda, formada por los monómeros de una familia de ADN satélite, puede ser recortada del gel y posteriormente amplificada e incluso secuenciada (Garrido-Ramos, 2017).

Además del estudio por electroforesis en gel de agarosa, esta técnica también es compatible con *Dot-blot* y *Southern blot* (Garrido-Ramos, 2017). El análisis con *Dot-blot* permite realizar una cuantificación aproximada de la cantidad de ADN satélite (de una familia concreta) presente en un genoma, mientras que el *Southern blot* permite obtener una estima de la variación de la secuencia dentro y entre especies. La tara de este procedimiento es su incapacidad para detectar ADN satélites que no presenten alguna diana de restricción en su secuencia (Garrido-Ramos, 2017).

El posterior advenimiento de las técnicas de secuenciación *NGS* en conjunto con nuevas herramientas bioinformáticas ha permitido un mejor estudio de este tipo de secuencias repetitivas (Garrido-Ramos, 2017).

RepeatExplorer es una herramienta diseñada originalmente para analizar los genomas de plantas superiores (*Computational resources were provided by the ELIXIR-CZ project (LM2015047), part of the international ELIXIR infrastructure*; Novák *et al.*, 2020). En cuanto a su funcionamiento, utiliza un algoritmo aglomerante de secuencias basado en gráficos que permite la identificación de repeticiones *de novo* sin necesidad de usar bases de datos de elementos ya conocidos. Es ideal para el análisis de lecturas *NGS* debido a que el algoritmo está diseñado para utilizar secuencias cortas del genoma muestreadas de forma aleatoria, pudiendo así determinar la abundancia de una familia de ADN satélite en base a la proporción de lecturas que presente. Además, presenta herramientas adicionales para ayudar en la clasificación de los elementos detectados (Novák *et al.*, 2013).

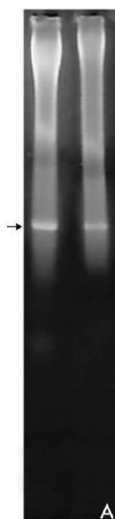


Figura 4: fragmento de la figura 1 de Garrido-Ramos (2017). En ella se observa una banda claramente marcada formada por los monómeros de una familia de ADN satélite sobre el *smear* del resto de ADN genómico.

La evolución cromosómica de la especie *Iberolacerta monticola* no estaba muy bien caracterizada. Es por este motivo por lo que ha sido el foco de diversos estudios. Durante su tesis doctoral, la Dra. Verónica Rojo Oróns (2015) investigó el ADN satélite de esta especie y de otras del mismo género, mediante el método basado en la digestión parcial con enzimas de restricción, que le permitió detectar dos familias, denominadas *HindIII* y *TaqI*, varios de cuyos representantes en el genoma consiguió secuenciar (Rojo Oróns, 2015).

Ella observó que esta especie presentaba 2 subfamilias de *HindIII*, las subfamilias I y II, y una única subfamilia de *TaqI* (Rojo Oróns, 2015). En la Tabla 1 se observa como la subfamilia I es 7.5 veces más frecuente que la II (ver también Giovannotti *et al.*, 2014, Rojo *et al.*, 2015 y Rojo Oróns, 2015). Cabe destacar el bajo número de secuencias analizadas en *I. monticola* (34 para *HindIII* y 10 para *TaqI*) a pesar de haberse realizado un gran esfuerzo experimental.

Species	Subfamily	HindIII			Subfamily	TaqI		
		n	Repeat length	Nucleotide diversity (π)		n	Repeat length	Nucleotide diversity (π)
<i>I. monticola</i>		34		0.0151 \pm 0.0018		10		0.0600 \pm 0.0089
	HI	30	171	0.0142 \pm 0.0023	TI	10	171 - 188	0.0600 \pm 0.0089
	HII	4	170	0.0177 \pm 0.0060				
<i>I. galani</i>		31		0.0331 \pm 0.0040		16		0.0489 \pm 0.0001
	HI	23	171	0.0148 \pm 0.0019	TI	16	186 - 188	0.0489 \pm 0.0001
	HII	8	169 - 170	0.0211 \pm 0.0082				
<i>I. martinezricai</i>		33		0.0151 \pm 0.0018		7		0.0541 \pm 0.0103
	HI	33	171 - 172	0.0151 \pm 0.0018	TI	7	187 - 188	0.0541 \pm 0.0103
<i>I. cyreni</i>		40		0.0356 \pm 0.0037		9		0.0406 \pm 0.0001
	HI	7		0.0180 \pm 0.0030	TI	9	186 - 187	0.0406 \pm 0.0001
	HII	33	161 - 171	0.0240 \pm 0.0029				
<i>I. horvathi</i>		12		0.0116 \pm 0.0028		33		0.1218 \pm 0.0079
	HI	12	171	0.0116 \pm 0.0028	TI	31	167 - 191	0.1184 \pm 0.0083
					TII	2	189 - 191	0.0699 \pm 0.0349
<i>I. aurelioi</i>		25		0.0396 \pm 0.0034		20		0.0976 \pm 0.0086
	HI	14	171	0.0290 \pm 0.0048	TI	1	187	
	HII	11	170	0.0262 \pm 0.0026	TII	19	177 - 188	0.0908 \pm 0.0074
<i>I. aranica</i>		22		0.0355 \pm 0.0043		34		0.1209 \pm 0.0070
	HI	7	151 - 171	0.0265 \pm 0.0055	TI	14	175 - 190	0.1082 \pm 0.0126
	HII	15	170	0.0164 \pm 0.0028	TII	20	177 - 190	0.0960 \pm 0.0059
<i>I. bonnali</i>		35		0.0491 \pm 0.0050		22		0.1204 \pm 0.0096
	HI	17	171	0.0257 \pm 0.0027	TI	17	155 - 188	0.1060 \pm 0.0102
	HII	15	169 - 170	0.0230 \pm 0.0076	TII	5	177 - 190	0.0983 \pm 0.0156
	HIII	3	171	0.0195 \pm 0.0033				
	HI	143		0.0241 \pm 0.0015	TI	105		0.1342 \pm 0.0060
	HII	53		0.0230 \pm 0.0018	TII	46		0.0961 \pm 0.0044
	HIII	36		0.0254 \pm 0.0029				
	TOTAL	232		0.0539 \pm 0.0020	TOTAL	151		0.1567 \pm 0.0038

Tabla 1: fragmento de la tabla 1 de la tesis doctoral de Verónica Rojo Oróns (2015). En ella se puede observar que *I. monticola* presenta una relación 30:4 entre sus subfamilias I y II (HI y HII) y que solo presenta a la subfamilia TI de *TaqI*.

2. Objetivos

En este TFM se ha estudiado, mediante el uso de la herramienta bioinformática *RepeatExplorer2*, si los resultados obtenidos con las técnicas clásicas de restricción parcial del ADN y secuenciación de una pequeña muestra aleatoria de productos de PCR (por cuanto se refiere a la abundancia relativa de familias y subfamilias) se ven confirmados utilizando *NGS* (de baja y media cobertura) y *RepeatExplorer2* (Novák *et al.*, 2013).

3. Materiales y métodos

Secuenciación del genoma de *Iberolacerta monticola*

Se encargó a la empresa AllGenetics la secuenciación del genoma completo de un macho y de una hembra de *Iberolacerta monticola*. La extracción de ADN fue llevada a cabo mediante el

uso de un *NZY Tissue gADN Isolation kit* (NZYTech), según las instrucciones del fabricante. Posteriormente se comprobó la integridad del ADN mediante un gel de agarosa al 2% teñido con *GreenSafe* (NZYTech; visualización con luz UV). Tras la cuantificación del ADN (*Qubit dsADN HS Assay Kit* de Thermo Fisher Scientific), se generaron librerías *shotgun* genómicas con un *TruSeq Nano ADN library Prep kit* (Illumina).

Tras la realización de las librerías, estas fueron cuantificadas (*Qubit dsADN HS Assay Kit* de Thermo Fisher Scientific) y secuenciadas en un *Illumina HiSeq PE150* (Figura 5), produciéndose lecturas *forward* (R1) y *reverse* (R2).

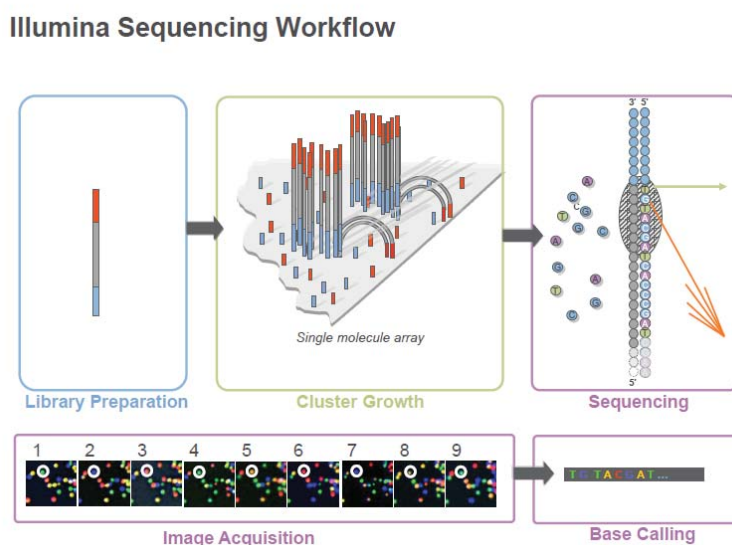


Figura 5: esquema de la línea de trabajo de una secuenciación NGS en *Illumina* (Dr. Naveira, comunicación personal). Se muestran los pasos de preparación de la librería (fragmentos genómicos con adaptadores ligados), amplificación por *bridge-PCR* y formación de clústers de secuencias y, por último, secuenciación por síntesis.

De esta especie, se realizaron 2 secuenciaciones independientes con coberturas distintas: en una de ellas, se secuenció el genoma de la hembra con una cobertura 26x, y en otra el genoma del macho, a cobertura 0.44x. El propósito de esto fue doble: poder analizar diferencias en la presencia de diferentes microsatélites entre los machos y las hembras y saber si realizar una secuenciación a una mayor cobertura aportaba algún beneficio asociado. A la hora de analizar diferencias entre los sexos, hay que tener en cuenta que, en esta especie, la hembra es el sexo heterogamético (WZ).

Análisis de la calidad de las lecturas por *FASTQC*

Tras la obtención de los ficheros *FASTQ*, se realizó un análisis de la calidad de las lecturas mediante *FASTQC*. Los resultados obtenidos para las lecturas R1 y R2 de macho y hembra están en el siguiente enlace de OneDrive. Las lecturas obtenidas de las hembra pasaron los criterios de selección: ninguna contiene secuencias con baja calidad (aunque baja un poco en la región 3' de las lecturas obtenidas con R2), además de tener una misma proporción de A, T, C y G a lo largo de las lecturas y un %GC normal. En cuanto a las de macho, ambos tipos de lecturas presentan algunas secuencias adaptadoras en su región 5', con el consiguiente sesgo en el porcentaje de GC en ese extremo. Debido a la naturaleza del análisis que estamos realizando, este tipo de secuencias no afectan a nuestros resultados. Se puede acceder a los 4 informes *FASTQC* obtenidos a través de este enlace de *OneDrive*:

https://udcgal-my.sharepoint.com/:f/g/personal/inaki_lopez_udc_es/EugUomf9PftDqAzIRrZS8hcBxZL92v6mI37LdQvDmp4OJw?e=1EqJdS

Pre-tratamiento

Una vez analizada la calidad de las lecturas, se llevó a cabo un pre-tratamiento de estas. Primero, se convirtió cada uno de los archivos *FASTQ* en *FASTA* (R1 y R2 de hembra y de macho). El nombre de estos archivos es: *sub1000_R1.fas* y *sub1000_R2.fas* para las lecturas de la hembra y *LA_01_R1.fas* y *LA_01_R2.fas* para las lecturas del macho.

Una vez convertidos, comprobé que se hubiesen convertido correctamente analizando el número de líneas que tenía cada archivo: los de hembra tenían 395717584 líneas (equivalente a 197858792 lecturas) y los de macho, 8229026 líneas (equivalente a 4114513 secuencias). Debido al elevado tamaño de los archivos R1 y R2 de la hembra, estos fueron fraccionados al azar, reduciéndose así las lecturas de 197858792 a 1000000. En cuanto a los archivos R1 y R2 del macho, estos no fueron reducidos debido a su menor tamaño. Al finalizar todo este proceso, chequeé que las lecturas no se hubiesen desaparejado entre cada pareja de archivos y las subí a *Galaxy*, una plataforma de acceso libre para la investigación computacional con uso intensivo de datos. Permite a los investigadores sin experiencia ejecutar diversos flujos de trabajo de análisis en sus datos, además de poder compartir sus resultados y permitir que otros repitan los mismos análisis (ELIXIR, s.f.; <https://elixir-europe.org/communities/galaxy>).

FASTA read affixer y *FASTA interlacer*

Una vez subidos a *Galaxy* los archivos FASTA R1 y R2 del macho y de la hembra, utilicé *FASTA read affixer* para añadir el prefijo “Ibero_sub1000” (hembra) o “LA_01” (macho) y el sufijo R1 o R2, en función de qué grupo de secuencias es (el resto de valores se dejaron por defecto). Después, procedí a entrelazar sus secuencias con *FASTA interlacer* con el fin de obtener un único archivo FASTA entrelazado para cada sexo (uso de valores por defecto).

Además, se utilizó también para realizar un análisis comparativo con las lecturas de los 2 sexos con el procedimiento indicado en Novák *et al.* (2020), páginas 20 a 21.

RepeatExplorer2 y TAREAN

Tras la confección de los archivos entrelazados para los dos sexos, se procedió a su análisis mediante la herramienta *RepeatExplorer2* de *Galaxy*. Se utilizaron los valores por defecto. Tras la finalización de *RepeatExplorer2*, se utilizaron principalmente los resultados obtenidos por *TAREAN* (*Tandem repeat analysis*) y *Repeat annotation summary*.

Además, se utilizó también *RepeatExplorer2* para realizar un análisis comparativo con las lecturas de los 2 sexos (valores por defecto).

EGassembler y UGENE

Para la identificación del clúster 1 (hembra), tuve que generar una secuencia consenso yo mismo, tomando los cóntigos generados por *RepeatExplorer2* y ensamblándolos con ayuda de la herramienta bioinformática *EGassembler* (<https://www.genome.jp/tools/egassembler/>; algoritmo CAP3). Únicamente utilicé su modo *One-Click Assembly*, dejando sus parámetros por defecto.

Una vez obtenido el ensamblaje, busqué la región del ensamblaje con la mayor cobertura y alineé sus cóntigos mediante el algoritmo *ClustalO* (parámetros por defecto) en *UGENE*. Terminado esto, solo quedó copiar la secuencia consenso generada por el programa, seleccionar la región de esta que presentaba una mayor cobertura y utilizarla como *query* en *BLASTn* (parámetros por defecto).

BLASTn

Con el fin de averiguar qué clústers contenían las secuencias de las familias *HindIII* y *TaqI*, se realizaron *BLASTn* con las secuencias consenso obtenidas en *TAREAN* para cada clúster (parámetros por defecto y búsqueda *MegaBLAST* a no ser que se especifique lo contrario).

También se utilizó en varias ocasiones a mayores para identificar a otros clústers de interés (como por ejemplo, al clúster de *SINE POM-99-like*).

Además, se hicieron varios *BLASTn* locales con el fin de analizar qué proporción de lecturas del clúster *HindIII* pertenecían a las subfamilias I y II de ese microsatélite. Para estos *BLASTn* locales se usaron como base de datos los cóntigos generados por *RepeatExplorer2* del clúster de *HindIII* de hembra y de macho. Se utilizaron las secuencias consenso de las subfamilias I y II como *query*, seleccionado exclusivamente la región comprendida entre las coordenadas 20 y 56, ya que es en la que más diferencias hay entre las dos subfamilias (Figura 6).



Figura 6: fragmento de la figura 1 de Giovannotti *et al.* (2014). En ella se muestran las secuencias consenso para diferentes variantes de la familia *HindIII* detectadas en varias especies del género *Iberolacerta*, entre ellas, *I. monticola* (IMO). La secuencia utilizada para el *BLASTn* local de la subfamilia I (_sfi) fue la que está señalada por un rectángulo negro. La secuencia utilizada para la subfamilia II está señalada por un rectángulo rojo (los puntos representan *matches* con la primera secuencia).

4. Resultados

El siguiente enlace permite acceder a la totalidad de los resultados obtenidos en *Galaxy*: <https://repeatexplorer-elixir.cerit-sc.cz/galaxy/u/horaci19-udc/h/workshop-repeatexplorer-2019>

Mis resultados están de la entrada 15 (esta incluida) en adelante.

Resultados generales

RepeatExplorer2 tomó muestras aleatorias de las lecturas que le fueron aportadas para su análisis de secuencias repetitivas. Usó 1979058 lecturas de hembra y 841791 lecturas de macho de sus respectivos archivos entrelazados. Estas lecturas fueron analizadas y organizadas en grupos (clústers) en base a sus similitudes entre ellas, y a su vez, estos grupos formaron otros grupos más grandes (superclústers) en función de las similitudes que presentaban entre ellos. Aquellas lecturas para las que no había otras similares formaron un grupo de *singlets* (Novak *et al.*, 2013).

En cuanto a la hembra, *RepeatExplorer2* generó 127233 superclústers (127240 clústers) con 744602 lecturas (22% del total de lecturas), dejando 1234456 lecturas como *singlets* (Figura 7). Los resultados obtenidos pueden visualizarse en la entrada 37: *RepeatExplorer2 – HTML report from data 27*.

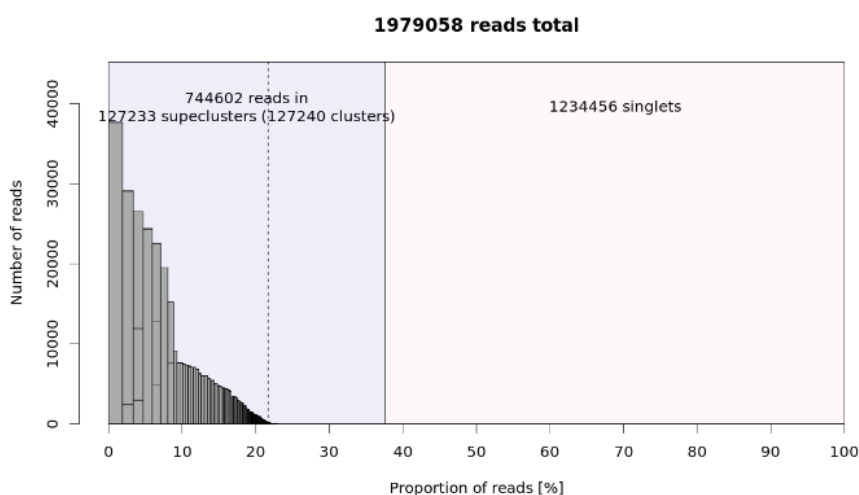


Figura 7: obtenida del informe de hembra de *I. monticola* en *RepeatExplorer2*. Representa el número de lecturas frente a la proporción de lecturas totales para cada superclúster.

El *Repeat annotation summary* mostró que el tipo de elemento repetitivo identificado más frecuente en nuestras lecturas son los LINES (Figura 8).

	Proportion[%]	Nsuperclusters	Nclusters	Nreads
Unclassified_repeat (conflicting evidences)	1.47	1	2	29082
--rDNA	0	0	0	0
--45S_rDNA	0	0	0	0
--18S_rDNA	0	0	0	0
--25S_rDNA	0	0	0	0
'--5.8S_rDNA	0	0	0	0
'--5S_rDNA	0	0	0	0
--satellite	1.77	6	6	35043
'--mobile_element	0	0	0	0
--Class_I	0	0	0	0
--SINE	0	0	0	0
'--LTR	0.15	1	1	2919
--Retrovirus	0	0	0	0
'--Ty3_gypsy	1.44	6	7	28422
'--Ty1_copia	0.48	5	5	9524
'--Bel-Pao	0	0	0	0
'--Penelope	0.32	1	1	6292
'--LINE	6.14	18	23	121601
'--DIRS	0	0	0	0
'--Class_II	0	0	0	0
--Subclass_1	0	0	0	0
'--TIR	0	0	0	0
'--MITE	0	0	0	0
'--Subclass_2	0	0	0	0
'--Maverick	0	0	0	0
'--Helitron	0	0	0	0
organelle	0	0	0	0
--plastid	0	0	0	0
'--mitochondria	0	0	0	0
Unclassified repeat (No evidence)	9.96	90	90	197200
contamination	0	0	0	0

Figura 8: Repeat annotation summary generado por *RepeatExplorer2* con las lecturas de hembra. Los **LINEs** forman un 6.14% de las lecturas totales.

A continuación, se muestra en la Figura 9 un gráfico circular que muestra con mayor claridad las proporciones mostradas en el *Repeat annotation summary*.

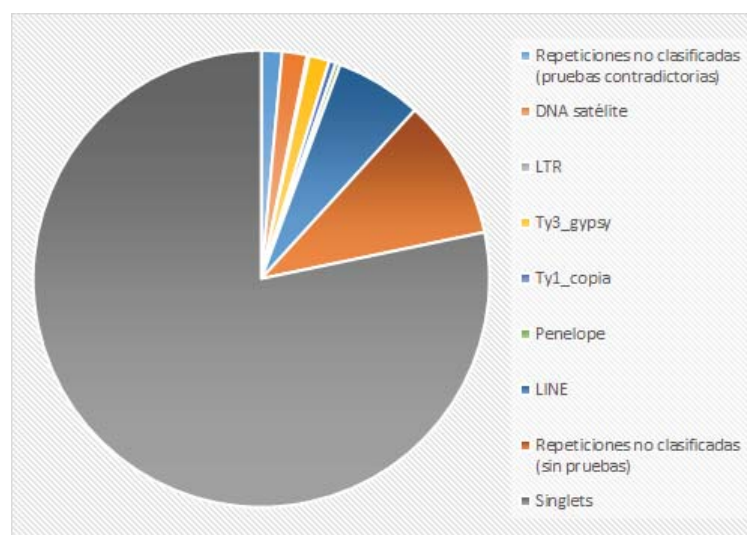


Figura 9: gráfico circular con las proporciones del *Repeat annotation summary* de hembra.

Por otro lado, 374162 lecturas (20% del total de lecturas) de macho fueron organizadas en 87608 superclústers (87614 clústers), quedando 467629 lecturas como *singlets* (Figura 10). Los

resultados obtenidos pueden visualizarse en la entrada 54: *RepeatExplorer2 – HTML report from data 44*.

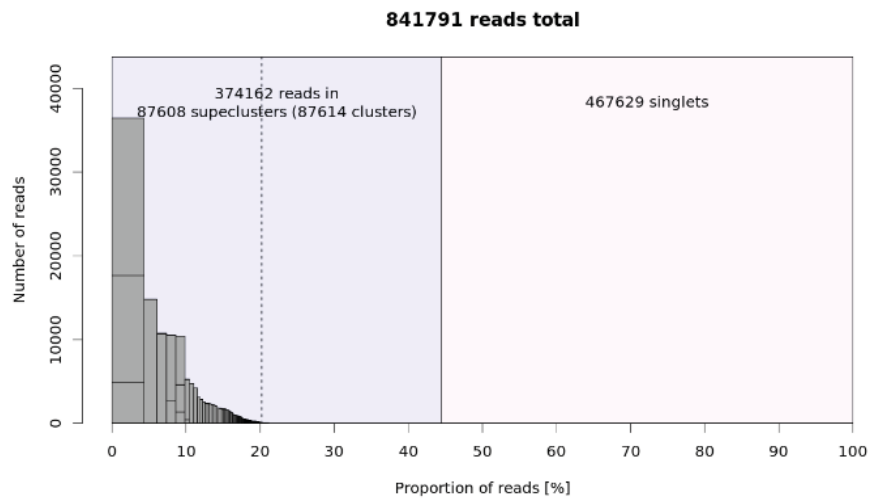


Figura 10: obtenida del informe de macho de *I. monticola* en *RepeatExplorer2*. Representa el número de lecturas frente a la proporción de lecturas totales para cada superclúster.

En cuanto a su *Repeat annotation summary*, en este caso se puede observar como de nuevo los LINES son los elementos repetitivos presentes en un mayor número de copias en nuestras lecturas (Figura 11). Además, *RepeatExplorer2* ha detectado la presencia de un transposón de ADN, concretamente un Helitron (0.08% de las lecturas), ausente en las lecturas de hembras.

Repeat annotation summary

This table summarizes the automatic annotations of superclusters that should be verified and manually corrected if necessary. Thus, the table should not be used as the final output of the analysis without critical evaluation.

	Proportion[%]	Nsuperclusters	Nclusters	Nreads
Unclassified_repeat (conflicting evidences)	0.62	1	2	5248
--rDNA	0	0	0	0
--45S_rDNA	0	0	0	0
--18S_rDNA	0	0	0	0
--25S_rDNA	0	0	0	0
--5.8S_rDNA	0	0	0	0
--5S_rDNA	0	0	0	0
--satellite	0.6	3	3	5025
--mobile_element	0	0	0	0
--Class_I	0	0	0	0
--SINE	0	0	0	0
--LTR	0	0	0	0
--Ty1_copia	0.32	3	3	2725
--Ale	0	0	0	0
--Alesia	0	0	0	0
--Angela	0	0	0	0
--Bianca	0	0	0	0
--Bryco	0	0	0	0
--Lyco	0	0	0	0
--Gymco-III	0	0	0	0
--Gymco-I	0	0	0	0
--Gymco-II	0	0	0	0
--Ikeros	0	0	0	0
--Ivana	0	0	0	0
--Gymco-IV	0	0	0	0
--Osser	0	0	0	0
--SIRE	0	0	0	0
--TAR	0	0	0	0
--Tork	0	0	0	0
--Ty1-outgroup	0	0	0	0
--Ty3_gypsy	0.86	5	5	7269
--non-chromovirus	0	0	0	0
--non-chromo-outgroup	0	0	0	0
--Phygy	0	0	0	0
--Seigy	0	0	0	0
--OTA	0	0	0	0
--Athila	0	0	0	0
--Tat	0	0	0	0
--TatI	0	0	0	0
--TatII	0	0	0	0
--TatIII	0	0	0	0
--Ogre	0	0	0	0
--Retand	0	0	0	0
--chromovirus	0.22	1	1	1850
--Chlamyvir	0	0	0	0
--Tcn1	0	0	0	0
--chromo-outgroup	0	0	0	0
--CRM	0	0	0	0
--Galadriel	0	0	0	0
--Tekay	0	0	0	0
--Reina	0	0	0	0
--chromo-unclass	0	0	0	0
--pararetrovirus	0	0	0	0
--DIRS	0	0	0	0
--Penelope	0	0	0	0
--LINE	3.24	7	10	27280
--Class_II	0	0	0	0
--Subclass_1	0	0	0	0
--TIR	0	0	0	0
--MITE	0	0	0	0
--EnSpm_CACTA	0	0	0	0
--hAT	0	0	0	0
--Kolobok	0	0	0	0
--Merlin	0	0	0	0
--MuDR_Mutator	0	0	0	0
--Novosib	0	0	0	0
--P	0	0	0	0
--PIF_Harbinger	0	0	0	0
--PiggyBac	0	0	0	0
--Sola1	0	0	0	0
--Sola2	0	0	0	0
--Tcl1_Mariner	0	0	0	0
--Subclass_2	0	0	0	0
--Helitron	0.08	1	1	692
organelle	0	0	0	0
--plastid	0	0	0	0
--mitochondria	0.14	3	3	1159
Unclassified repeat (No evidence)	14.12	108	110	118861
contamination	0	0	0	0

Figura 11: Repeat annotation summary generado por *RepeatExplorer2* con las lecturas de macho. En este caso, los **LINES** forman el 3.24% de las lecturas totales.

La Figura 12 muestra un gráfico circular que muestra las proporciones mostradas en el *Repeat annotation summary*.

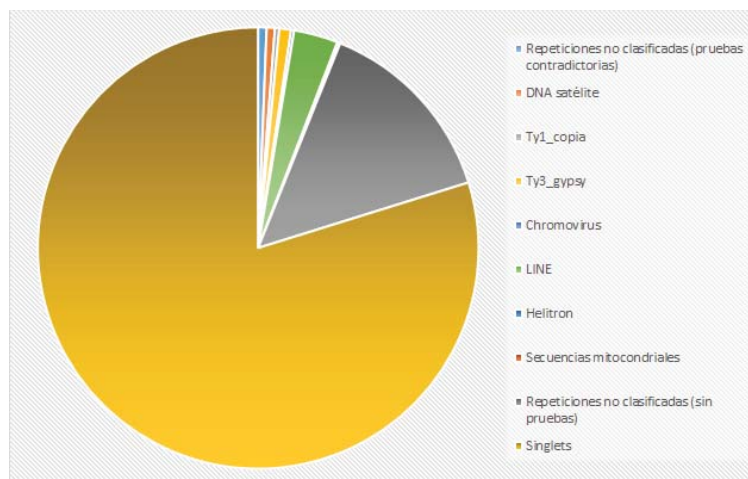


Figura 12: gráfico circular con las proporciones del *Repeat annotation summary* de macho.

Por último, en el estudio comparativo, *RepeatExplorer2* utilizó 1000000 lecturas, generando 71577 superclústers (71582 clústers) con 374596 lecturas (20% del total de lecturas). Clasificó a 625404 lecturas como *singlets* (Figura 13). Los resultados obtenidos pueden visualizarse en la entrada 63: *RepeatExplorer2 – HTML report from data 60*.

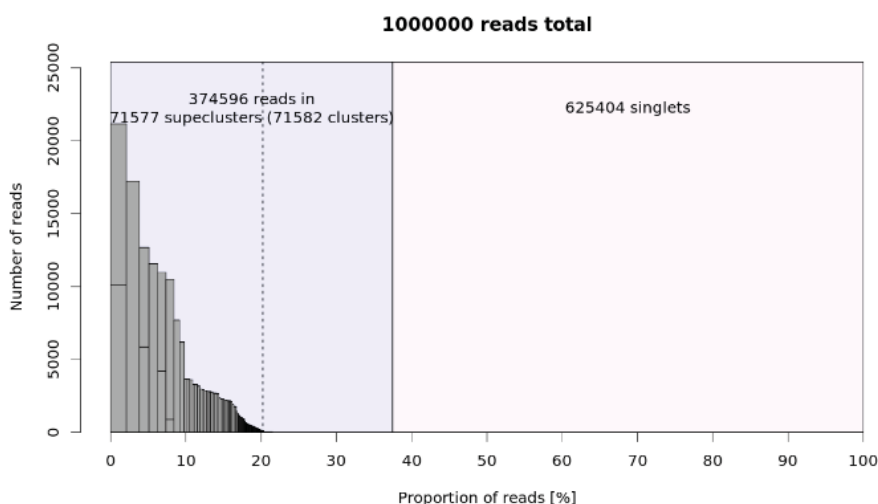


Figura 13: obtenida del informe del estudio comparativo de los sexos de *I. monticola* en *RepeatExplorer2*. Representa el número de lecturas frente a la proporción de lecturas totales para cada superclúster.

Casos de estudio

mtADN de la especie

Este caso de estudio deriva de los resultados obtenidos para *I. monticola* hembra (*Tandem Repeat Analysis* en 37: *RepeatExplorer2 – HTML report from data 27*). Lo primero que me

llamó la atención tras observar por primera vez la tabla de clústers producida por TAREAN fue la presencia de una secuencia de ADN satélite mucho más larga que el resto de secuencias presentes en las otras tablas. El clúster 27, al que pertenecía esta secuencia consenso, presentaba un gráfico con una forma muy característica, típica de los clústers de mtADN (Figura 14; Dr. Naveira, comunicación personal). Tras una búsqueda en *BLASTn* (valores por defecto), se concluyó que la secuencia pertenecía al mtADN de la especie (Figura 15). Esta secuencia fue detectada por *RepeatExplorer2* debido a su elevado número de copias por célula. Este clúster contiene un 0.26% de las lecturas totales analizadas en hembra, y por motivos que desconocemos no fue anotado automáticamente como mitocondrial por *RepeatExplorer2*, aunque sí lo fue en el análisis realizado a partir de las lecturas de macho.

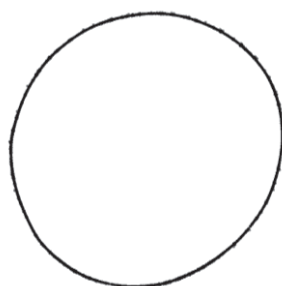


Figura 14: diagrama del clúster 27, correspondiente al mtADN de *Iberolacerta monticola*. Cada vértice representa a una lectura.

Descriptions		Graphic Summary	Alignments	Taxonomy
Sequences producing significant alignments		Download	New Select columns	Show 100
select all 100 sequences selected		GenBank	Graphics	Distance tree of results
Description	Scientific Name	Max Score	Total Score	Query Cover
<input checked="" type="checkbox"/> Darevskia parvula mitochondrion complete genome	Darevskia parvula	8425	15349	99%
<input checked="" type="checkbox"/> Darevskia portschinskii mitochondrion complete genome	Darevskia portsc...	8423	15271	99%
<input checked="" type="checkbox"/> Darevskia nudis mitochondrion complete genome	Darevskia nudis	8419	15321	99%
<input checked="" type="checkbox"/> Darevskia valentini mitochondrion complete genome	Darevskia valentini	8397	15317	99%
<input checked="" type="checkbox"/> Darevskia mixta mitochondrion complete genome	Darevskia mixta	8392	15420	99%
<input checked="" type="checkbox"/> Darevskia daghestanica mitochondrion complete genome	Darevskia daghe...	8384	15412	99%
<input checked="" type="checkbox"/> Darevskia dahli mitochondrion complete genome	Darevskia dahli	8375	15431	99%
<input checked="" type="checkbox"/> Darevskia armeniaca mitochondrion complete genome	Darevskia arme...	8375	15420	99%
<input checked="" type="checkbox"/> Podarcis muralis mitochondrion complete genome	Podarcis muralis	8349	14792	99%
<input checked="" type="checkbox"/> Podarcis muralis mitochondrion complete genome	Podarcis muralis	8349	14792	99%
<input checked="" type="checkbox"/> Darevskia caucasica mitochondrion complete genome	Darevskia cauca...	8298	14665	94%
<input checked="" type="checkbox"/> Darevskia saxicola mitochondrion complete genome	Darevskia saxicola	8296	15105	99%
<input checked="" type="checkbox"/> Darevskia raddei mitochondrion complete genome	Darevskia raddei	8283	15127	99%
<input checked="" type="checkbox"/> Darevskia braueri mitochondrion complete genome	Darevskia braueri	8264	15042	99%
<input checked="" type="checkbox"/> Darevskia unisexualis mitochondrion complete genome	Darevskia unise...	8233	15024	99%
<input checked="" type="checkbox"/> Darevskia clarkorum mitochondrion complete genome	Darevskia clarko...	8198	14216	92%
Alburnus nannumitatus voucher MVI 8/C-731 mitochondrion complete genome	Alburnus nann...	8170	13465	97%

Figura 15: resultado de *BLASTn* obtenido usando como *query* la secuencia consenso del clúster 27 de hembra. Se puede observar como presenta un porcentaje de identidad muy elevado (99%) con los mtADNs de otras especies de Lacertidae.

Estudio de las subfamilias I y II de HindIII

El análisis de estas subfamilias se realizó a través del reporte de *TAREAN* para la hembra (*Tandem Repeat Analysis en 37: RepeatExplorer2 – HTML report from data 27*). Como ya he mencionado con anterioridad, se sabe que la familia *HindIII* está presente como 2 subfamilias (I y II) en esta especie y que la *TaqI*, como una (Giovannotti *et al.*, 2014).

Al observar el gráfico del clúster 2 (*HindIII*), se puede observar dos núcleos diferentes de vértices, uno con una cantidad mayor de vértices que el otro (Figura 16). En base a esta observación se puede hacer la suposición de que el núcleo más poblado representa a la subfamilia I y el que es más reducido, a la subfamilia II. Esta suposición correspondería con la información previa sobre la proporción de cada subfamilia presente en el genoma de esta especie: la subfamilia I es 7.5 veces (30 a 4) más frecuente que la II (Rojo Oróns, 2015).

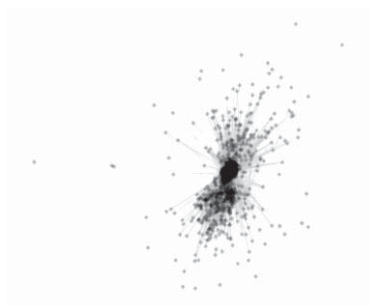


Figura 16: gráfico del clúster 2 de hembra (familia *HindIII*). Este clúster contiene un 1.30% de todas las lecturas de hembra analizadas. Se puede observar un núcleo secundario por debajo del núcleo principal.

En cuanto al clúster 4 (*TaqI*), se puede observar en su gráfico como claramente hay un núcleo de lecturas muy similares, concordando con el conocimiento previo de esta familia de ADN satélite (solo una subfamilia en este genoma; Figura 17).



Figura 17: gráfico del clúster 4 de hembra (familia *TaqI*). Este clúster contiene un 0.99% de todas las lecturas de hembra analizadas. Se puede observar un único núcleo rodeado de secuencias similares.

Tras varias búsquedas en *BLASTn*, encontré que los clústers 2 (Figura 18) y 4 (Figura 19) correspondían a las familias de ADN satélites *HindIII* y *TaqI*.

Descriptions	Graphic Summary	Alignments	Taxonomy
--------------	-----------------	------------	----------

Sequences producing significant alignments

Download

New

 Select columns Show 100

?

☒ select all 100 sequences selected

GenBankGraphicsDistance tree of resultsNewMSA Viewer

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	Iberolacerta martinezricai satellite DNA locus HindIII, clone IMR-174a	Iberolacerta martinezricai	257	316	99%	2e-64	100.00%	171	LN879209.1
<input checked="" type="checkbox"/>	Iberolacerta galani satellite DNA locus HindIII, clone IGA-168c	Iberolacerta galani	257	316	100%	2e-64	100.00%	171	LN879160.1
<input checked="" type="checkbox"/>	Iberolacerta aurelioi satellite DNA locus HindIII, clone IAU-305	Iberolacerta aurelioi	257	318	100%	2e-64	100.00%	171	LN879084.1
<input checked="" type="checkbox"/>	Iberolacerta martinezricai satellite DNA locus HindIII, clone IMR-279c	Iberolacerta martinezricai	252	312	100%	7e-63	99.28%	171	LN879223.1
<input checked="" type="checkbox"/>	Iberolacerta martinezricai satellite DNA locus HindIII, clone IMR-245	Iberolacerta martinezricai	252	252	83%	7e-63	99.28%	171	LN879216.1
<input checked="" type="checkbox"/>	Iberolacerta martinezricai satellite DNA locus HindIII, clone IMR-170a	Iberolacerta martinezricai	252	252	83%	7e-63	99.28%	171	LN879204.1
<input checked="" type="checkbox"/>	Iberolacerta monticola satellite DNA locus HindIII, clone IMO-146c	Iberolacerta monticola	252	312	100%	7e-63	99.28%	171	LN879192.1
<input checked="" type="checkbox"/>	Iberolacerta monticola satellite DNA locus HindIII, clone IMO-20a	Iberolacerta monticola	252	312	100%	7e-63	99.28%	171	LN879181.1
<input checked="" type="checkbox"/>	Iberolacerta horvathi satellite DNA locus HindIII, clone IHO-66	Iberolacerta horvathi	252	312	100%	7e-63	99.28%	171	LN879179.1
<input checked="" type="checkbox"/>	Iberolacerta horvathi satellite DNA locus HindIII, clone IHO-61a	Iberolacerta horvathi	252	252	83%	7e-63	99.28%	171	LN879174.1
<input checked="" type="checkbox"/>	Iberolacerta galani satellite DNA locus HindIII, clone IGA-192a	Iberolacerta galani	252	312	100%	7e-63	99.28%	171	LN879166.1
<input checked="" type="checkbox"/>	Iberolacerta galani satellite DNA locus HindIII, clone IGA-169	Iberolacerta galani	252	312	100%	7e-63	99.28%	171	LN879161.1
<input checked="" type="checkbox"/>	Iberolacerta galani satellite DNA locus HindIII, clone IGA-160c	Iberolacerta galani	252	312	100%	7e-63	99.28%	171	LN879153.1
<input checked="" type="checkbox"/>	Iberolacerta martinezricai clone 17 satellite HindIII sequence	Iberolacerta martinezricai	252	252	83%	7e-63	99.28%	171	KF453675.1
<input checked="" type="checkbox"/>	Iberolacerta monticola clone 15 satellite HindIII sequence	Iberolacerta monticola	252	312	100%	7e-63	99.28%	171	KF453667.1
<input checked="" type="checkbox"/>	Iberolacerta galani satellite DNA locus HindIII, clone IGA-160a	Iberolacerta galani	252	312	100%	7e-63	99.28%	171	LN879151.1
<input checked="" type="checkbox"/>	Iberolacerta bonnali satellite DNA locus HindIII, clone IBN-114b	Iberolacerta bonnali	252	312	100%	7e-63	99.28%	171	LN879149.1

Feedback

Figura 18: fragmento del informe obtenido con la secuencia consenso del clúster 2 de hembra (familia *HindIII*) en *BLASTn*.

Descriptions	Graphic Summary	Alignments	Taxonomy					
Sequences producing significant alignments								
Download ▼ New Select columns ▼ Show 100 ▼ ?								
<input checked="" type="checkbox"/> select all 100 sequences selected GenBank Graphics Distance tree of results New MSA Viewer								
Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> Iberolacerta galani clone 13 satellite TaqI sequence	Iberolacerta galani	268	342	100%	8e-68	99.32%	187	KF453695.1
<input checked="" type="checkbox"/> Iberolacerta galani clone 42 satellite TaqI sequence	Iberolacerta galani	263	336	100%	4e-66	98.65%	187	KF453706.1
<input checked="" type="checkbox"/> Iberolacerta bonnali satellite DNA locus TaqI, clone IBN-25a	Iberolacerta bonnali	257	330	100%	2e-64	97.97%	187	LN879281.1
<input checked="" type="checkbox"/> Iberolacerta galani clone 31 satellite TaqI sequence	Iberolacerta galani	257	257	80%	2e-64	97.97%	187	KF453704.1
<input checked="" type="checkbox"/> Iberolacerta galani clone 17 satellite TaqI sequence	Iberolacerta galani	257	330	100%	2e-64	97.97%	187	KF453698.1
<input checked="" type="checkbox"/> Iberolacerta galani clone 16 satellite TaqI sequence	Iberolacerta galani	257	330	100%	2e-64	97.99%	188	KF453697.1
<input checked="" type="checkbox"/> Iberolacerta monticola clone 24 satellite TaqI sequence	Iberolacerta monticola	255	323	100%	6e-64	97.97%	186	KF453711.1
<input checked="" type="checkbox"/> Iberolacerta monticola clone 13 satellite TaqI sequence	Iberolacerta monticola	252	252	80%	8e-63	97.30%	187	KF453707.1
<input checked="" type="checkbox"/> Iberolacerta galani clone 1 satellite TaqI sequence	Iberolacerta galani	252	312	100%	8e-63	97.30%	186	KF453693.1
<input checked="" type="checkbox"/> Iberolacerta galani isolate 1 clone 32 satellite TaqI sequence	Iberolacerta galani	252	325	100%	8e-63	97.30%	187	KF453692.1
<input checked="" type="checkbox"/> Iberolacerta galani clone 22 satellite TaqI sequence	Iberolacerta galani	250	323	100%	3e-62	97.30%	186	KF453701.1
<input checked="" type="checkbox"/> Iberolacerta martinezricai clone 21 satellite TaqI sequence	Iberolacerta martinezricai	246	319	100%	4e-61	96.62%	187	KF453719.1
<input checked="" type="checkbox"/> Iberolacerta martinezricai clone 1 satellite TaqI sequence	Iberolacerta martinezricai	246	246	80%	4e-61	96.62%	187	KF453717.1
<input checked="" type="checkbox"/> Iberolacerta galani clone 1B satellite TaqI sequence	Iberolacerta galani	246	316	98%	4e-61	96.62%	187	KF453691.1
<input checked="" type="checkbox"/> Iberolacerta cyreni clone 8 satellite TaqI sequence	Iberolacerta cyreni	246	319	100%	4e-61	96.62%	187	KF453689.1
<input checked="" type="checkbox"/> Iberolacerta cyreni clone 4 satellite TaqI sequence	Iberolacerta cyreni	244	318	100%	1e-60	96.62%	187	KF453688.1
<input checked="" type="checkbox"/> Iberolacerta bonnali satellite DNA locus TaqI, clone IBN-161	Iberolacerta bonnali	241	314	100%	2e-59	95.95%	187	KF453687.1


 Feedback

Figura 19: fragmento del informe obtenido con la secuencia consenso del clúster 4 de hembra (familia *HindIII*) en *BLASTn*.

Para confirmar la presencia de las 2 subfamilias en las muestras de hembra, se realizó una búsqueda *BLASTn* local. Para estos *BLASTn* locales se usaron como base de datos los cóntigos generados por *RepeatExplorer2* del clúster de *HindIII*.

Basándome en el valor de *score* como medida del grado de identidad entre la *query* y los *hits*, la búsqueda confirmó la presencia en un mayor número la subfamilia I frente a la II (Figura 20 y Figura 21, respectivamente).

Los informes obtenidos muestran como en efecto, aparece un mayor número de veces la subfamilia I que la subfamilia II, concretamente en una relación 92:12. La subfamilia I es 7.66 veces más frecuente que la subfamilia II.

```

BLASTN 2.11.0+

Reference:
Zheng Zhang, Scott Schwartz, Lukas Wagner, and Webb Miller (2000),
"A greedy algorithm for aligning DNA sequences", J Comput Biol 2000;
7(1-2):203-14.

Database: contigs_2.fasta
        202 sequences; 51,927 total letters

Query= consenso_subfamilia_I
Length=37

Sequences producing significant alignments:

CL2Contig201      69.4  1e-15
CL2Contig197      69.4  1e-15
CL2Contig193      69.4  1e-15
CL2Contig192      69.4  1e-15
CL2Contig189      69.4  1e-15
CL2Contig188      69.4  1e-15
CL2Contig184      69.4  1e-15
CL2Contig180      69.4  1e-15
CL2Contig179      69.4  1e-15
CL2Contig178      69.4  1e-15
CL2Contig177      69.4  1e-15
CL2Contig170      69.4  1e-15
CL2Contig165      69.4  1e-15
CL2Contig151      69.4  1e-15
CL2Contig147      69.4  1e-15
CL2Contig143      69.4  1e-15
CL2Contig141      69.4  1e-15
CL2Contig140      69.4  1e-15
CL2Contig131      69.4  1e-15
CL2Contig124      69.4  1e-15
CL2Contig118      69.4  1e-15

```

Figura 20: fragmento del informe del *BLASTn* local de la secuencia consenso (20-56) de la subfamilia *HindIII* I usando como base de datos los cóntigos del clúster 2.


```

BLASTN 2.11.0+

Reference:
Zheng Zhang, Scott Schwartz, Lukas Wagner, and Webb Miller (2000),
"A greedy algorithm for aligning DNA sequences", J Comput Biol 2000;
7(1-2):203-14.

Database: contigs_2.fasta
        202 sequences; 51,927 total letters

Query= consenso_subfamilia_II

Length=37

Sequences producing significant alignments:

CL2Contig95          69.4    1e-15
CL2Contig31          69.4    1e-15
CL2Contig168         67.6    5e-15
CL2Contig152         67.6    5e-15
CL2Contig145         67.6    5e-15
CL2Contig121         67.6    5e-15
CL2Contig108         67.6    5e-15
CL2Contig84          67.6    5e-15
CL2Contig104         63.9    7e-14
CL2Contig190         58.4    3e-12
CL2Contig158         58.4    3e-12
CL2Contig70          58.4    3e-12

```

Figura 21: fragmento del informe del *BLASTn* local de la secuencia consenso (20-56) de la subfamilia *HindIII* II usando como base de datos los cóntigos del clúster 2.

También cabe mencionar que a su vez se analizaron las lecturas de macho mediante *RepeatExplorer2* (*Tandem Repeat Analysis* en 54: *RepeatExplorer2 – HTML report from data 44*). También se detectaron ambas familias: *HindIII* en el clúster 8 (Figura 22) y *TaqI* en el clúster 9 (Figura 23). Su identificación se llevó a cabo mediante una búsqueda de sus secuencias consenso en *BLASTn* (parámetros por defecto; Figura 24 -*HindIII*- y Figura 25 -*TaqI*-). En ambos clúster solo se aprecia un único núcleo de lecturas rodeado de un número significativamente menor de secuencias en comparación con los clústers equivalentes de hembra.

A pesar de ello, tras realizar un *BLASTn* local utilizando como *query* a las secuencias consenso de las subfamilias I y II de *HindIII*, y, como base de datos, los cóntigos generados por *RepeatExplorer2* para el clúster 8, se detectó la presencia de secuencias de ambas subfamilias. Al contrario que en la hembra, la subfamilia II es más frecuente que la I, con una relación 11:33 (la subfamilia II es 3 veces más frecuente que la I).

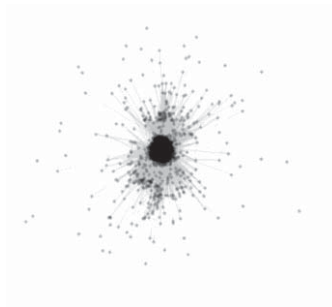


Figura 22: gráfico del clúster 8 de macho, representante de la familia *HindIII*. Este clúster contiene un 0.57% de todas las lecturas de macho analizadas. En este caso, se observa un único núcleo central de lecturas muy similares.

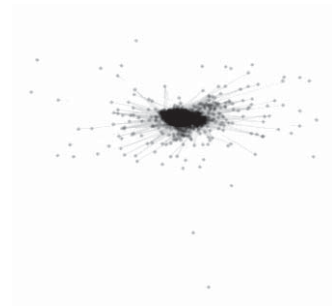


Figura 23: gráfico del clúster 9 de macho, representante de la familia *TaqI*. Este clúster contiene un 0.57% de todas las lecturas de macho analizadas. Tal y como se esperaba, se observa un único núcleo central de lecturas muy similares al igual que en hembras.

Descriptions	Graphic Summary	Alignments	Taxonomy
--------------	-----------------	------------	----------

Sequences producing significant alignments

Download

New

 Select columns Show 100

☒ select all 100 sequences selected

GenBankGraphicsDistance tree of resultsNewMSA Viewer

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	Iberolacerta aranica satellite DNA locus HindIII . clone IAR-116	Iberolacerta aranica	246	310	100%	3e-61	100.00%	170	LN879054.1
<input checked="" type="checkbox"/>	Iberolacerta aranica satellite DNA locus HindIII . clone IAR-103	Iberolacerta aranica	246	246	80%	3e-61	100.00%	170	LN879047.1
<input checked="" type="checkbox"/>	Iberolacerta aranica satellite DNA locus HindIII . clone IAR-108a	Iberolacerta aranica	246	316	100%	3e-61	100.00%	170	LN879048.1
<input checked="" type="checkbox"/>	Iberolacerta bonnali satellite DNA locus HindIII . clone IBN-148b	Iberolacerta bonnali	241	310	100%	2e-59	99.25%	170	LN879116.1
<input checked="" type="checkbox"/>	Iberolacerta aurelioi satellite DNA locus HindIII . clone IAU-32a	Iberolacerta aurelioi	241	241	80%	2e-59	99.25%	170	LN879064.1
<input checked="" type="checkbox"/>	Iberolacerta aranica satellite DNA locus HindIII . clone IAR-169	Iberolacerta aranica	241	310	100%	2e-59	99.25%	170	LN879056.1
<input checked="" type="checkbox"/>	Iberolacerta aranica satellite DNA locus HindIII . clone IAR-113	Iberolacerta aranica	241	310	100%	2e-59	99.25%	170	LN879052.1
<input checked="" type="checkbox"/>	Iberolacerta aranica satellite DNA locus HindIII . clone IAR-109	Iberolacerta aranica	241	310	100%	2e-59	99.25%	170	LN879050.1
<input checked="" type="checkbox"/>	Iberolacerta galani clone 45 satellite HindIII sequence	Iberolacerta galani	241	310	100%	2e-59	99.25%	170	KF453658.1
<input checked="" type="checkbox"/>	Iberolacerta aranica satellite DNA locus HindIII . clone IAR-112	Iberolacerta aranica	241	310	100%	2e-59	99.25%	170	LN879051.1
<input checked="" type="checkbox"/>	Iberolacerta galani clone 33 satellite HindIII sequence	Iberolacerta galani	241	241	80%	2e-59	99.25%	170	KF453653.1
<input checked="" type="checkbox"/>	Iberolacerta bonnali satellite DNA locus HindIII . clone IBN-148a	Iberolacerta bonnali	241	310	100%	2e-59	99.25%	170	LN879115.1
<input checked="" type="checkbox"/>	Iberolacerta bonnali satellite DNA locus HindIII . clone IBN-140a	Iberolacerta bonnali	239	239	79%	5e-59	99.24%	170	LN879111.1
<input checked="" type="checkbox"/>	Iberolacerta bonnali satellite DNA locus HindIII . clone IBN-54a	Iberolacerta bonnali	237	237	78%	2e-58	99.24%	170	LN879090.1
<input checked="" type="checkbox"/>	Iberolacerta aurelioi satellite DNA locus HindIII . clone IAU-30	Iberolacerta aurelioi	235	305	100%	7e-58	98.50%	170	LN879063.1
<input checked="" type="checkbox"/>	Iberolacerta aranica satellite DNA locus HindIII . clone IAR-178	Iberolacerta aranica	235	235	80%	7e-58	98.50%	170	LN879062.1
<input checked="" type="checkbox"/>	Iberolacerta bonnali satellite DNA locus HindIII . clone IBN-233	Iberolacerta bonnali	235	305	100%	7e-58	98.50%	170	LN879061.1

Feedback

Figura 24: fragmento del informe obtenido con la secuencia consenso del clúster 8 de macho (familia *HindIII*) en *BLASTn*.

Descriptions									
Sequences producing significant alignments									
<input checked="" type="checkbox"/> select all 100 sequences selected Download New Select columns Show 100 ?									
GenBank Graphics Distance tree of results New MSA Viewer									
Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession	
<input checked="" type="checkbox"/> Iberolacerta galani clone 42 satellite TaqI sequence	Iberolacerta galani	246	342	100%	4e-61	99.26%	187	KF453706.1	
<input checked="" type="checkbox"/> Iberolacerta galani clone 31 satellite TaqI sequence	Iberolacerta galani	241	330	100%	2e-59	98.53%	187	KF453704.1	
<input checked="" type="checkbox"/> Iberolacerta galani clone 17 satellite TaqI sequence	Iberolacerta galani	241	336	100%	2e-59	98.53%	187	KF453698.1	
<input checked="" type="checkbox"/> Iberolacerta galani clone 16 satellite TaqI sequence	Iberolacerta galani	241	336	100%	2e-59	98.54%	188	KF453697.1	
<input checked="" type="checkbox"/> Iberolacerta galani clone 13 satellite TaqI sequence	Iberolacerta galani	241	336	100%	2e-59	98.53%	187	KF453695.1	
<input checked="" type="checkbox"/> Iberolacerta monticola clone 24 satellite TaqI sequence	Iberolacerta monticola	239	329	100%	6e-59	98.53%	186	KF453711.1	
<input checked="" type="checkbox"/> Iberolacerta martinezricai clone 1 satellite TaqI sequence	Iberolacerta martinezricai	235	235	73%	8e-58	97.79%	187	KF453717.1	
<input checked="" type="checkbox"/> Iberolacerta monticola clone 13 satellite TaqI sequence	Iberolacerta monticola	235	235	73%	8e-58	97.79%	187	KF453707.1	
<input checked="" type="checkbox"/> Iberolacerta galani clone 1 satellite TaqI sequence	Iberolacerta galani	235	318	100%	8e-58	97.79%	186	KF453693.1	
<input checked="" type="checkbox"/> Iberolacerta bonnali satellite DNA locus TaqI clone IBN-160	Iberolacerta bonnali	230	230	73%	4e-56	97.06%	187	LN879300.1	
<input checked="" type="checkbox"/> Iberolacerta bonnali satellite DNA locus TaqI clone IBN-25a	Iberolacerta bonnali	230	325	100%	4e-56	97.06%	187	LN879281.1	
<input checked="" type="checkbox"/> Iberolacerta martinezricai clone 21 satellite TaqI sequence	Iberolacerta martinezricai	230	325	100%	4e-56	97.06%	187	KF453719.1	
<input checked="" type="checkbox"/> Iberolacerta galani clone 33 satellite TaqI sequence	Iberolacerta galani	230	319	98%	4e-56	97.06%	187	KF453705.1	
<input checked="" type="checkbox"/> Iberolacerta galani clone 1B satellite TaqI sequence	Iberolacerta galani	230	308	100%	4e-56	97.06%	187	KF453691.1	
<input checked="" type="checkbox"/> Iberolacerta cyreni clone 8 satellite TaqI sequence	Iberolacerta cyreni	230	325	100%	4e-56	97.06%	187	KF453689.1	
<input checked="" type="checkbox"/> Iberolacerta cyreni clone 4 satellite TaqI sequence	Iberolacerta cyreni	228	323	100%	1e-55	97.06%	187		
<input checked="" type="checkbox"/> Iberolacerta cyreni clone 20 satellite TaqI sequence	Iberolacerta cyreni	228	318	99%	1e-55	97.04%	187		

Figura 25: fragmento del informe obtenido con la secuencia consenso del clúster 9 de macho (familia *TaqI*) en *BLASTn*.

SINE POM-99-like

El clúster 1 femenino (*Tandem Repeat Analysis* en 37: *RepeatExplorer2 – HTML report from data 27*) es uno de los clústers que presentan una mayor abundancia con respecto al número total de lecturas (1.9%). Además, presenta un gráfico con una estructura peculiar: un núcleo de secuencias muy similares entre sí rodeado de algunas pocas secuencias a una distancia considerable (Figura 26).



Figura 26: gráfico del clúster 1 de hembra. Este clúster, que conforma un porcentaje importante (1.90%) de las lecturas totales de hembra, está formado principalmente por lecturas muy similares entre sí.

Tras una búsqueda de la zona con mayor cobertura de su secuencia consenso en *BLASTn*, el algoritmo detectó una coincidencia altamente significativa (86.73% de porcentaje de identidad y un valor de E de $5 \cdot 10^{-57}$) con el *SINE POM-99*, de *Podarcis muralis*, una especie también de la familia Lacertidae, por lo que he denominado a este elemento *SINE POM-99-like*.

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	PREDICTED: Zootoca vivipara uncharacterized LOC118091733 (LOC118091733), transcript variant X2, nc...	Zootoca vivipara	276	276	95%	8e-70	86.13%	5233	XR_004693430.1
<input checked="" type="checkbox"/>	PREDICTED: Zootoca vivipara A-kinase anchoring protein 13 (AKAP13), transcript variant X9, mRNA	Zootoca vivipara	270	270	96%	4e-68	85.51%	14860	XM_035130710.1
<input checked="" type="checkbox"/>	PREDICTED: Zootoca vivipara zinc-binding protein A33-like (LOC118081059), transcript variant X2, mRNA	Zootoca vivipara	243	243	96%	8e-60	83.70%	2616	XM_035107195.1
<input checked="" type="checkbox"/>	PREDICTED: Zootoca vivipara zinc-binding protein A33-like (LOC118081059), transcript variant X1, mRNA	Zootoca vivipara	243	243	96%	8e-60	83.70%	2631	XM_035107194.1
<input checked="" type="checkbox"/>	PREDICTED: Zootoca vivipara E3 ubiquitin-protein ligase TRIM7-like (LOC118077420), mRNA	Zootoca vivipara	243	243	96%	8e-60	83.70%	2297	XM_035101039.1
<input checked="" type="checkbox"/>	Podarcis muralis clone POM-99 SINE, complete sequence	Podarcis muralis	233	233	76%	5e-57	86.73%	352	DQ023371.1
<input checked="" type="checkbox"/>	Darevskia raddei clone dar43 Squam-1 SINE repeat sequence	Darevskia raddei	231	231	81%	2e-56	85.59%	426	DQ393697.1
<input checked="" type="checkbox"/>	Podarcis muralis clone POM-80 SINE, complete sequence	Podarcis muralis	228	228	76%	2e-55	86.43%	347	DQ023363.1
<input checked="" type="checkbox"/>	Podarcis muralis clone POM-75 SINE, complete sequence	Podarcis muralis	220	220	75%	4e-53	86.24%	357	DQ023360.1
<input checked="" type="checkbox"/>	PREDICTED: Zootoca vivipara zinc finger and BTB domain containing 46 (ZBTB46), transcript variant X3,...	Zootoca vivipara	219	219	96%	1e-52	82.31%	3646	XM_035122830.1
<input checked="" type="checkbox"/>	Podarcis muralis clone POM-21 SINE, complete sequence	Podarcis muralis	211	211	76%	2e-50	84.82%	355	DQ023344.1
<input checked="" type="checkbox"/>	PREDICTED: Zootoca vivipara potassium calcium-activated channel subfamily M alpha 1 (KCNMA1), trans...	Zootoca vivipara	209	209	96%	8e-50	81.52%	13179	XM_035137305.1
<input checked="" type="checkbox"/>	Darevskia raddei clone dar37 Squam-1 SINE repeat sequence	Darevskia raddei	209	209	91%	8e-50	82.02%	420	DQ393696.1
<input checked="" type="checkbox"/>	Podarcis muralis clone POM-113 SINE, complete sequence	Podarcis muralis	207	207	77%	3e-49	84.44%	344	DQ023379.1
<input checked="" type="checkbox"/>	PREDICTED: Zootoca vivipara WAS/WASL-interacting protein family member 1-like (LOC118093102), tran...	Zootoca vivipara	204	204	81%	4e-48	83.83%	5128	XR_004693598.1
<input checked="" type="checkbox"/>	PREDICTED: Zootoca vivipara WAS/WASL-interacting protein family member 1-like (LOC118093102), tran...	Zootoca vivipara	204	204	81%	4e-48	83.83%	4839	XM_035137305.1
<input checked="" type="checkbox"/>	Darevskia raddei strain EGE(1-73) SINE Squam1, complete sequence	Darevskia raddei	204	204	76%	4e-48	84.62%	386	KX000000000.1

Figura 27: fragmento del resultado de *BLASTn* de la secuencia consenso del clúster 1 de hembra. La secuencia consenso es un 86.73% idéntica a *SINE POM-99-like* de *Podarcis muralis*.

Análisis comparativo

Los resultados de este caso de estudio se encuentran en el informe *Tandem Repeat Analysis* en la entrada del historial 63: *RepeatExplorer2 – HTML report from data 60*, y se resumen en la Figura 28. Como puede verse en el gráfico inferior (*proportion of individual samples*), existe en general una buena correspondencia entre la abundancia de los distintos clústers en uno y otro sexo, con excepciones que merecen mención aparte.

En la figura 28 se puede observar la presencia de un único clúster (clúster 28) formado exclusivamente por lecturas presentes tan solo en hembras. Este resultado era de esperar debido a que esta especie presenta determinación sexual cromosómica ZW, siendo las hembras el sexo heterogamético. Por lo tanto, podría ser posible que ese clúster esté formado por secuencias presentes únicamente en el cromosoma W (Figura 29).

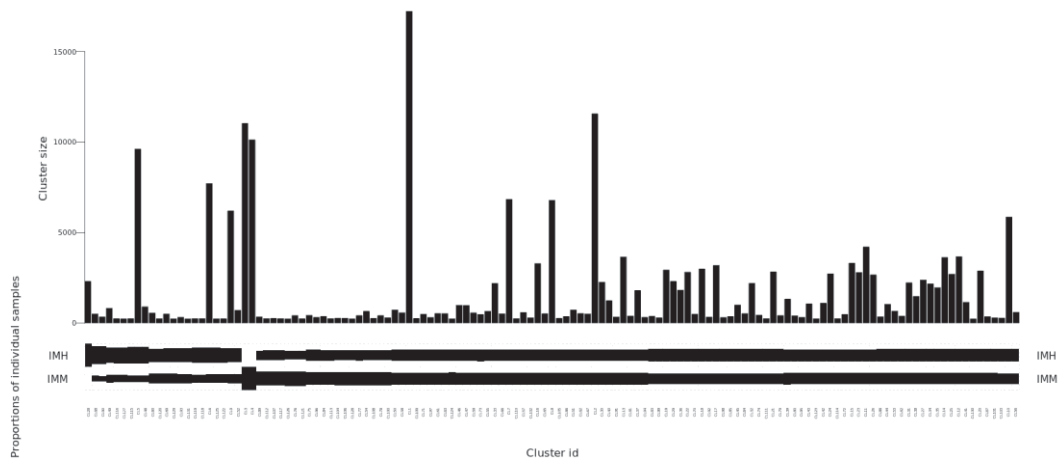


Figura 28: diagrama obtenido del análisis comparativo de las lecturas de hembra frente a las de macho de *I. monticola*. Representa el número de lecturas de los distintos clústers (arriba) y la proporciones de muestras individuales (abajo) en hembras (IMH) y machos (IMM).

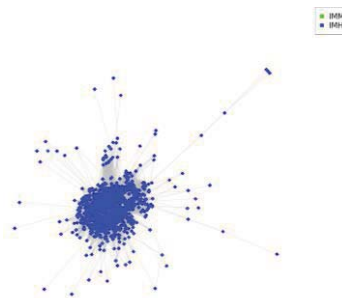


Figura 29: gráfico del clúster 28 del análisis comparativo. Este clúster contiene un 0.23% de todas las lecturas del análisis comparativo. Se puede observar como las lecturas que lo forman son exclusivamente de hembras. En azul, lecturas de hembra; en verde, lecturas de macho.

Tras una búsqueda en *BLASTn* (búsqueda *BLAST*) se pudo confirmar que este clúster representa a un ADN satélite poco conservado de la familia Lacertidae (Figura 30). Por desgracia, los investigadores que subieron estas secuencias a la base de datos del NCBI no incluyeron su localización cromosómica, pero todo indica que están ligadas al cromosoma W.

<div> <div>Descriptions</div> <div>Graphic Summary</div> <div>Alignments</div> <div>Taxonomy</div> </div>									
Sequences producing significant alignments									
<div> <div>Download</div> <div>New Select columns</div> <div>Show 100</div> </div>									
<div> <div>select all 58 sequences selected</div> <div> <div>GenBank</div> <div>Graphics</div> <div>Distance tree of results</div> <div>New MSA Viewer</div> </div> </div>									
	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	Lacerta derjugini satellite DNA	Darevskia derju...	78.8	78.8	82%	9e-11	75.41%	146	AJ252096.1
<input checked="" type="checkbox"/>	Lacerta armeniaca satellite DNA	Darevskia arme...	76.1	76.1	82%	1e-09	74.38%	146	AJ252095.1
<input checked="" type="checkbox"/>	Lacerta raddei satellite DNA	Darevskia raddei	76.1	76.1	82%	1e-09	74.38%	146	AJ252101.1
<input checked="" type="checkbox"/>	Lacerta rudis satellite DNA	Darevskia rudis	71.6	71.6	82%	1e-08	73.55%	146	AJ252103.1
<input checked="" type="checkbox"/>	Lacerta portschinskii satellite DNA	Darevskia ports...	71.6	71.6	82%	1e-08	73.55%	146	AJ252102.1
<input checked="" type="checkbox"/>	Lacerta dahli satellite DNA clone 1	Darevskia dahli	71.6	71.6	82%	1e-08	73.55%	146	AJ252093.1
<input checked="" type="checkbox"/>	Darevskia lindholmi clone CLsat1 satellite lin1-12, partial sequence	Darevskia lindho...	68.9	68.9	76%	2e-07	74.11%	146	AY256941.1
<input checked="" type="checkbox"/>	Darevskia lindholmi clone CLsat1 satellite lin1-6, partial sequence	Darevskia lindho...	68.9	68.9	70%	2e-07	75.96%	146	AY256935.1
<input checked="" type="checkbox"/>	Lacerta lindholmi satellite DNA clone 1	Darevskia lindho...	68.9	68.9	76%	2e-07	74.11%	146	AJ252097.1
<input checked="" type="checkbox"/>	Darevskia lindholmi clone CLsat1 satellite lin1-5, partial sequence	Darevskia lindho...	68.0	68.0	63%	2e-07	76.60%	146	AY256934.1
<input checked="" type="checkbox"/>	Darevskia rostombekovi clone 3#4ros1 satellite CLsat1 sequence	Darevskia rosto...	67.1	67.1	61%	6e-07	76.92%	146	AY262962.1
<input checked="" type="checkbox"/>	Darevskia mixta clone 1#54rmix1 satellite CLsat1 sequence	Darevskia mixta	66.2	66.2	78%	6e-07	73.91%	145	AY262988.1
<input checked="" type="checkbox"/>	Darevskia dryada clone #3dry1 satellite CLsat1 sequence	Darevskia dryada	64.4	64.4	62%	2e-06	76.09%	146	AY262973.1
<input checked="" type="checkbox"/>	Darevskia lindholmi clone CLsat1 satellite lin1-13, partial sequence	Darevskia lindho...	64.4	64.4	76%	2e-06	73.21%	146	AY256942.1
<input checked="" type="checkbox"/>	Darevskia lindholmi clone CLsat1 satellite lin1-4, partial sequence	Darevskia lindho...	64.4	64.4	76%	2e-06	73.21%	146	AY256933.1
<input checked="" type="checkbox"/>	Darevskia lindholmi clone CLsat1 satellite lin1-1, partial sequence	Darevskia lindho...	64.4	64.4	76%	2e-06	73.21%	146	AY256933.1

Figura 30: fragmento del informe obtenido con *BLASTn*. En él se muestran los *hits* obtenidos usando como *query* la secuencia consenso del clúster 28 del análisis comparativo. Los porcentajes de identidad obtenidos con estos alineamientos son relativamente bajos: el más elevado es de un 75.41% con un ADN satélite de *Lacerta derjugini*.

Además de los clústers exclusivos del sexo femenino, también se detectaron 2 clústers exclusivos de macho: los clústers 3 (Figura 31) y 4 (Figura 32). Estos clústers conforman el 2.2% de las lecturas totales: 1.1% el clúster 3 y 1.0%, el 4. Otro dato de especial relevancia es que comparten 9231 lecturas, resultando en una similitud entre ellos del 83.67%. Fueron identificados automáticamente por *RepeatExplorer2* como elementos LINE y tras una búsqueda en *BLASTn* (parámetros por defecto), el algoritmo detectó una coincidencia muy significativa (96.55% de porcentaje de identidad y un valor de E de $2 \cdot 10^{-30}$) con el LINE *Bov-B* de *Darevskia raddei*, una especie también de la familia Lacertidae. Por el momento, no encontramos explicación para la extraordinaria sobreabundancia de este elemento en el sexo masculino.



Figura 31: gráfico del clúster 3 del análisis comparativo. Este clúster contiene un 1.10% de todas las lecturas del análisis comparativo. El clúster está formado principalmente por lecturas de

macho, con alguna lectura de hembra dispar. En azul, lecturas de hembra; en verde, lecturas de macho.



Figura 32: gráfico del clúster 4 del análisis comparativo. Este clúster contiene un 1.00% de todas las lecturas del análisis comparativo. Similar al caso anterior, el clúster está formado principalmente por lecturas de macho, con alguna lectura de hembra dispar. En azul, lecturas de hembra; en verde, lecturas de macho.

El análisis comparativo de las lecturas de ambos sexos puso también de manifiesto los rasgos llamativos de otros clústers. Así, el clúster 111 presenta una estructura peculiar en la que se distinguen dos núcleos claramente definidos, unidos por unos cuantos vértices (Figura 33). Estos dos núcleos no corresponden uno a secuencias masculinas y el otro a femeninas, sino que se encuentran entremezcladas.

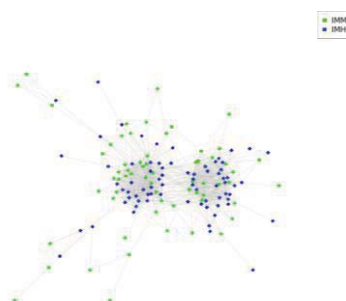


Figura 33: gráfico del clúster 111 del análisis comparativo. Este clúster contiene un 0.01% de todas las lecturas del análisis comparativo. Se puede observar como los dos núcleos claramente definidos (izquierda y derecha) están formado tanto por lecturas masculinas (vértices verdes) como femeninas (vértices azules).

Tras buscar la secuencia consenso del clúster en *BLASTn* (parámetros por defecto), encontré que parece estar compuesto por secuencias codificantes de estructuras similares a dedos de Zn (Figura 34).

Descriptions	Graphic Summary	Alignments	Taxonomy					
Sequences producing significant alignments								
Download New Select columns Show 100 ?								
<input checked="" type="checkbox"/> select all 100 sequences selected								
GenBank Graphics Distance tree of results New MSA Viewer								
Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> PREDICTED: Zootoca vivipara oocyte zinc finger protein XICOF6-like (LOC118093496). mRNA	Zootoca vivipara	139	614	100%	2e-29	96.43%	3165	XM_035132724.1
<input checked="" type="checkbox"/> PREDICTED: Lacerta agilis oocyte zinc finger protein XICOF6-like (LOC117042272). partial mRNA	Lacerta agilis	139	1016	100%	2e-29	96.43%	1881	XM_033141804.1
<input checked="" type="checkbox"/> PREDICTED: Podarcis muralis zinc finger protein 420-like (LOC114592588). mRNA	Podarcis muralis	139	1238	100%	2e-29	96.43%	2229	XM_028720860.1
<input checked="" type="checkbox"/> PREDICTED: Zootoca vivipara zinc finger protein 14-like (LOC118081192). transcript variant X2. mRNA	Zootoca vivipara	135	618	100%	3e-28	96.34%	3529	XM_035107498.1
<input checked="" type="checkbox"/> PREDICTED: Zootoca vivipara zinc finger protein 14-like (LOC118081192). transcript variant X1. mRNA	Zootoca vivipara	135	969	100%	3e-28	96.34%	5182	XM_035107497.1
<input checked="" type="checkbox"/> PREDICTED: Lacerta agilis zinc finger protein 180-like (LOC117042280). partial mRNA	Lacerta agilis	135	504	97%	3e-28	96.34%	1737	XM_033141814.1
<input checked="" type="checkbox"/> PREDICTED: Lacerta agilis uncharacterized LOC117042262 (LOC117042262). mRNA	Lacerta agilis	135	5462	100%	3e-28	96.34%	9306	XM_033141795.1
<input checked="" type="checkbox"/> PREDICTED: Lacerta agilis zinc finger protein 658B-like (LOC117042197). transcript variant X2. mRNA	Lacerta agilis	135	555	100%	3e-28	96.34%	2662	XM_033141719.1
<input checked="" type="checkbox"/> PREDICTED: Lacerta agilis zinc finger protein 658B-like (LOC117042197). transcript variant X1. mRNA	Lacerta agilis	135	555	100%	3e-28	96.34%	3424	XM_033141718.1
<input checked="" type="checkbox"/> PREDICTED: Podarcis muralis zinc finger protein 208-like (LOC114591186). mRNA	Podarcis muralis	135	1849	100%	3e-28	96.34%	6476	XM_028717854.1
<input checked="" type="checkbox"/> PREDICTED: Zootoca vivipara oocyte zinc finger protein XICOF6-like (LOC118080973). mRNA	Zootoca vivipara	134	639	100%	1e-27	95.24%	3450	XM_035107046.1
<input checked="" type="checkbox"/> PREDICTED: Lacerta agilis zinc finger protein 658B-like (LOC117042782). partial mRNA	Lacerta agilis	134	236	100%	1e-27	95.24%	1530	XM_033142428.1
<input checked="" type="checkbox"/> PREDICTED: Lacerta agilis zinc finger protein 420-like (LOC117042717). transcript variant X1. mRNA	Lacerta agilis	134	491	100%	1e-27	95.24%	2430	XM_033142336.1
<input checked="" type="checkbox"/> PREDICTED: Lacerta agilis zinc finger protein 208-like (LOC117042608). mRNA	Lacerta agilis	134	1222	100%	1e-27	95.24%	3754	XM_033142146.1
<input checked="" type="checkbox"/> PREDICTED: Lacerta agilis zinc finger protein 729-like (LOC117042597). partial mRNA	Lacerta agilis	134	897	100%	1e-27	95.24%	3186	XM_033142133.1
<input checked="" type="checkbox"/> PREDICTED: Lacerta agilis zinc finger protein 420-like (LOC117042196). transcript variant X2. mRNA	Lacerta agilis	134	397	100%	1e-27	95.24%	1806	XM_033142132.1
<input checked="" type="checkbox"/> PREDICTED: Lacerta agilis zinc finger protein 420-like (LOC117042196). transcript variant X1. mRNA	Lacerta agilis	134	582	100%	1e-27	95.24%	2876	XM_033142131.1


 Feedback

Figura 34: fragmento del informe obtenido con la secuencia consenso del clúster 111 del análisis comparativo en *BLASTn*.

Estas secuencias predichas son todas pertenecientes a especies de Lacertidae y presentan un porcentaje de identidad muy elevado con mi secuencia consenso del clúster 111 (máximo de 96.43%), señalando un posible grado de conservación muy elevado.

En cuanto a las familias de ADN satélites *HindIII* y *TaqI*, también forman clústers en el análisis comparativo. Estas 2 familias corresponderían a los clústers 6 (Figura 35) y 5 (Figura 36), respectivamente. Esto fue confirmado mediante la búsqueda de sus secuencias consenso en *BLASTn* (parámetros por defecto; Figura 37 y Figura 38, respectivamente).

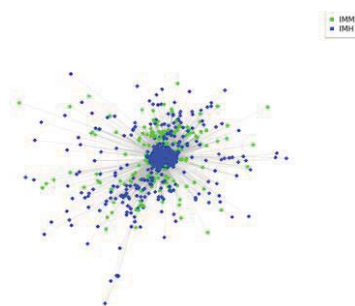


Figura 35: gráfico del clúster 6 del análisis comparativo. Este clúster contiene un 0.77% de todas las lecturas del análisis comparativo y corresponde con la familia *HindIII* de *I. monticola*. En azul, lecturas de hembra; en verde, lecturas de macho.

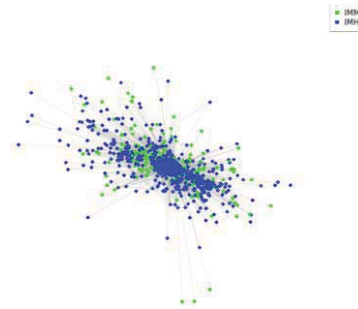


Figura 36: gráfico del clúster 5 del análisis comparativo. Este clúster contiene un 0.96% de todas las lecturas del análisis comparativo y corresponde con la familia *TaqI* de *I. monticola*. En azul, lecturas de hembra; en verde, lecturas de macho.

Descriptions									
Sequences producing significant alignments									
Download Select columns Show 100									
select all 100 sequences selected									
GenBank Graphics Distance tree of results MSA Viewer									
Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession	
<input checked="" type="checkbox"/> Iberolacerta martinezricai satellite DNA locus HindIII, clone IMR-174a	Iberolacerta martinezricai	257	316	99%	1e-64	100.00%	171	LN879209.1	
<input checked="" type="checkbox"/> Iberolacerta galani satellite DNA locus HindIII, clone IGA-168c	Iberolacerta galani	257	316	100%	1e-64	100.00%	171	LN879160.1	
<input checked="" type="checkbox"/> Iberolacerta aurelioi satellite DNA locus HindIII, clone IAU-305	Iberolacerta aurelioi	257	318	100%	1e-64	100.00%	171	LN879084.1	
<input checked="" type="checkbox"/> Iberolacerta martinezricai satellite DNA locus HindIII, clone IMR-279c	Iberolacerta martinezricai	252	312	100%	7e-63	99.28%	171	LN879223.1	
<input checked="" type="checkbox"/> Iberolacerta martinezricai satellite DNA locus HindIII, clone IMR-245	Iberolacerta martinezricai	252	252	83%	7e-63	99.28%	171	LN879216.1	
<input checked="" type="checkbox"/> Iberolacerta martinezricai satellite DNA locus HindIII, clone IMR-170a	Iberolacerta martinezricai	252	252	83%	7e-63	99.28%	171	LN879204.1	
<input checked="" type="checkbox"/> Iberolacerta monticola satellite DNA locus HindIII, clone IMO-146c	Iberolacerta monticola	252	312	100%	7e-63	99.28%	171	LN879192.1	
<input checked="" type="checkbox"/> Iberolacerta monticola satellite DNA locus HindIII, clone IMO-20a	Iberolacerta monticola	252	312	100%	7e-63	99.28%	171	LN879181.1	
<input checked="" type="checkbox"/> Iberolacerta horvathi satellite DNA locus HindIII, clone IHO-66	Iberolacerta horvathi	252	312	100%	7e-63	99.28%	171	LN879179.1	
<input checked="" type="checkbox"/> Iberolacerta horvathi satellite DNA locus HindIII, clone IHO-61a	Iberolacerta horvathi	252	252	83%	7e-63	99.28%	171	LN879174.1	
<input checked="" type="checkbox"/> Iberolacerta galani satellite DNA locus HindIII, clone IGA-192a	Iberolacerta galani	252	312	100%	7e-63	99.28%	171	LN879166.1	
<input checked="" type="checkbox"/> Iberolacerta galani satellite DNA locus HindIII, clone IGA-169	Iberolacerta galani	252	312	100%	7e-63	99.28%	171	LN879161.1	
<input checked="" type="checkbox"/> Iberolacerta galani satellite DNA locus HindIII, clone IGA-160c	Iberolacerta galani	252	312	100%	7e-63	99.28%	171	LN879153.1	
<input checked="" type="checkbox"/> Iberolacerta martinezricai clone 17 satellite HindIII sequence	Iberolacerta martinezricai	252	252	83%	7e-63	99.28%	171	KF453675.1	
<input checked="" type="checkbox"/> Iberolacerta monticola clone 15 satellite HindIII sequence	Iberolacerta monticola	252	312	100%	7e-63	99.28%	171	KF453667.1	
<input checked="" type="checkbox"/> Iberolacerta galani satellite DNA locus HindIII, clone IGA-160a	Iberolacerta galani	252	312	100%	7e-63	99.28%	171		
<input checked="" type="checkbox"/> Iberolacerta bonnali satellite DNA locus HindIII, clone IBN-114b	Iberolacerta bonnali	252	312	100%	7e-63	99.28%	171		

Figura 37: fragmento del informe obtenido con la secuencia consenso del clúster 6 del análisis comparativo en *BLASTn*. Presenta un 100% de identidad con las secuencias de la familia *HindIII* en otras especies de Lacertidae.

Descriptions									
Sequences producing significant alignments									
<input checked="" type="checkbox"/> select all 100 sequences selected GenBank Graphics Distance tree of results New MSA Viewer									
Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession	
<input checked="" type="checkbox"/> Iberolacerta galani clone 13 satellite TaqI sequence	Iberolacerta galani	252	342	100%	8e-63	99.28%	187	KF453695.1	
<input checked="" type="checkbox"/> Iberolacerta monticola clone 13 satellite TaqI sequence	Iberolacerta monticola	246	246	75%	4e-61	98.56%	187	KF453707.1	
<input checked="" type="checkbox"/> Iberolacerta galani clone 42 satellite TaqI sequence	Iberolacerta galani	246	336	100%	4e-61	98.56%	187	KF453706.1	
<input checked="" type="checkbox"/> Iberolacerta bonnali satellite DNA locus TaqI clone IBN-25a	Iberolacerta bonnali	241	330	100%	2e-59	97.84%	187	LN879281.1	
<input checked="" type="checkbox"/> Iberolacerta galani clone 31 satellite TaqI sequence	Iberolacerta galani	241	325	100%	2e-59	97.84%	187	KF453704.1	
<input checked="" type="checkbox"/> Iberolacerta galani clone 17 satellite TaqI sequence	Iberolacerta galani	241	330	100%	2e-59	97.84%	187	KF453698.1	
<input checked="" type="checkbox"/> Iberolacerta galani clone 16 satellite TaqI sequence	Iberolacerta galani	241	330	100%	2e-59	97.86%	188	KF453697.1	
<input checked="" type="checkbox"/> Iberolacerta galani clone 18 satellite TaqI sequence	Iberolacerta galani	241	318	99%	2e-59	97.84%	187	KF453691.1	
<input checked="" type="checkbox"/> Iberolacerta monticola clone 24 satellite TaqI sequence	Iberolacerta monticola	239	323	100%	6e-59	97.84%	186	KF453711.1	
<input checked="" type="checkbox"/> Iberolacerta martinezricai clone 1 satellite TaqI sequence	Iberolacerta martinezricai	235	235	75%	8e-58	97.12%	187	KF453717.1	
<input checked="" type="checkbox"/> Iberolacerta galani clone 1 satellite TaqI sequence	Iberolacerta galani	235	312	100%	8e-58	97.12%	186	KF453693.1	
<input checked="" type="checkbox"/> Iberolacerta galani isolate 1 clone 32 satellite TaqI sequence	Iberolacerta galani	235	325	100%	8e-58	97.12%	187	KF453692.1	
<input checked="" type="checkbox"/> Iberolacerta galani clone 22 satellite TaqI sequence	Iberolacerta galani	233	323	100%	3e-57	97.12%	186	KF453701.1	
<input checked="" type="checkbox"/> Iberolacerta bonnali satellite DNA locus TaqI clone IBN-160	Iberolacerta bonnali	230	230	75%	4e-56	96.40%	187	LN879300.1	
<input checked="" type="checkbox"/> Iberolacerta martinezricai clone 21 satellite TaqI sequence	Iberolacerta martinezricai	230	319	100%	4e-56	96.40%	187	KF453719.1	
<input checked="" type="checkbox"/> Iberolacerta cyreni clone 8 satellite TaqI sequence	Iberolacerta cyreni	230	319	100%	4e-56	96.40%	187	KF453719.1	
<input checked="" type="checkbox"/> Iberolacerta galani clone 33 satellite TaqI sequence	Iberolacerta galani	228	318	99%	1e-55	96.38%	187	KF453719.1	

Figura 38: fragmento del informe obtenido con la secuencia consenso del clúster 5 del análisis comparativo en *BLASTn*. Presenta un 99.28% de identidad con las secuencias de la familia *TaqI* en otras especies de *Lacertidae*.

5. Discusión

El *Repeat annotation summary* muestra una serie de características curiosas en cuanto a la composición de secuencias repetitivas entre los genomas estudiados. Por un lado, muestra como tanto en hembras como en machos los elementos repetitivos más frecuentes (que *RepeatExplorer2* haya podido identificar) son los LINEs. Este resultado era de esperar, ya que se sabe que los LINEs forman un porcentaje muy importante del genoma de muchos organismos eucariotas (e. g., el genoma humano; Miller & Therman, 2001). Por otro lado, también se puede observar la aparición de secuencias de un transposón de ADN en machos: el transposón *Helitron*. Estos transposones de ADN son capaces de duplicarse mediante replicación de círculo rodante (Kapitonov & Jurka, 2007). Han sido detectados tanto en animales como en plantas, mostrando ser una importante herramienta evolutiva debido a su capacidad para capturar genes del genoma hospedador (Kapitonov & Jurka, 2007). *A priori* se puede considerar que esta secuencia es exclusiva de este sexo debido a que la secuenciación en el macho fue a una cobertura mucho menor a la llevada a cabo en la hembra, por lo que sería extraño que esta secuencia no hubiese aparecido ni una vez entre las lecturas de la hembra si estuviese en su genoma. Su presencia exclusiva en las lecturas masculinas es desconcertante debido a que, al ser el sexo homogamético, se esperaría que todas sus secuencias estuviesen presentes también en la hembra ZW. Otro tanto cabe decir de los clusters 3 y 4 del análisis comparativo, cuya distribución entre los sexos muestra un acusadísimo sesgo.

Con respecto a la obtención de la secuencia completa del mtADN de *Iberolacerta monticola*, este fue un resultado inesperado debido a que no se diseñó el estudio con el fin de detectar a este tipo de secuencias. Aún así, tiene sentido que se detectase en un análisis de secuencias repetitivas una secuencia como esa debido a su naturaleza multicopia. El conocimiento de esta secuencia es muy importante a la hora de estudiar la historia evolutiva de una especie. Crochet *et al.* (2004) utilizó al gen mitocondrial del citocromo B (*cytb*) con el fin de construir un árbol filogenético que recogiese las relaciones evolutivas entre las especies del género *Iberolacerta*. Sus resultados mostraron que esta especie es un grupo hermano de *I. cantabrica* y que a su vez, estas dos forman un clado hermano de *I. martinezricai* (Crochet *et al.*, 2004).

En cuanto a los resultados relacionados con las familias de ADN satélites *HindIII* y *TaqI*, ha quedado confirmada la capacidad de *RepeatExplorer2* para reconocer y formar clústers que contengan las secuencias de familias de ADN satélite descubiertas mediante métodos tradicionales con enzimas de restricción. Además, los clústers que contienen a estas secuencias presentan formas reconocibles, pudiendo hacer una identificación rápida *in situ* de los clústers simplemente observando la forma de su gráfico. Tanto es así, que incluso se pueden indentificar “sub-clústers” en el interior de ciertos clústers como fue en el caso del gráfico del clúster 8 femenino, donde se puede distinguir la presencia de las 2 subfamilias de *HindIII*.

Un resultado que llama la atención es la gran diferencia observada entre la proporción subfamilia I: subfamilia II de la familia *HindIII* al comparar mis valores en el macho con los obtenidos por Verónica Rojo Oróns (2015). Debido a que el cálculo se realizó con los resultados obtenidos con lecturas de *NGS* (aleatorias), esta diferencia no debería ser causada por sesgos en la proporción de lecturas.

Además de ADN satélites, *RepeatExplorer2* también es capaz de encontrar secuencias repetitivas dispersas, como al elemento transponible *SINE POM-99-like*. EL clúster al que pertenece esta secuencia se mostró como un núcleo de lecturas extremadamente similares rodeado de algunas lecturas dispares. Este resultado es de esperar debido a la naturaleza de este elemento genético. Debido a que es un SINE (elemento retrotransposable no viral), esta secuencia carece de genes que codifiquen una retrotranscriptasa y a una endonucleasa y por lo tanto no es capaz de replicarse y moverse por el genoma por su cuenta (Piskurek *et al.*, 2006). Para poder hacerlo requiere de algún LINE que contenga una copia funcional de estos genes (Piskurek *et al.*, 2006).

Todo este proceso implica que la expansión de un SINE por el genoma es un proceso unidireccional: el SINE puede insertarse en el genoma, pero no puede eliminarse con facilidad (Piskurek *et al.*, 2006). Es por ello que la presencia de una de estas secuencias siempre se considera un carácter derivado cuando se compara con otro genoma que no contenga tal inserción (Piskurek *et al.*, 2006). Sabiendo esto, se puede razonar entonces que la presencia de esta secuencia en *I. monticola* y en *P. muralis* conlleva que este elemento transponible se debió de insertar en el genoma de sus ancestros antes de la separación de los 2 linajes. Por otro lado, su gran nivel de conservación podría ser debido a que en ocasiones, ciertos SINEs pueden actuar como secuencias reguladoras de la transcripción génica. Un ejemplo de esto lo presentaron Crepaldi *et al.* (2013), que demostraron como genes neuronales regulados por actividad requieren de un SINE cercano al que se le pueda unir TFIIC y así pueda llevarse a cabo su transcripción. Como resultado de esto, estos elementos transponibles ahora estarían sometidos a la selección natural, que evitaría que se produjesen grandes modificaciones en su secuencia. Cabe destacar que *RepeatExplorer2* no fue capaz de identificar directamente a este SINE, poniendo en manifiesto su naturaleza original como *software* de análisis de genomas vegetales.

Por último, los resultados del análisis comparativo demuestran la presencia de clústers compuestos por secuencias exclusivas de cada sexo. El clúster 28, formado por un ADN satélite poco conservado, fue el único clúster compuesto exclusivamente por lecturas de hembra. Este resultado era de esperar, ya que tiene sentido que existan ADN satélites presentes únicamente en hembras debido a la presencia del cromosoma W en su genoma. Ya en 1980, Singh *et al.* (1980) observaron la presencia de ADN satélite exclusivo de hembras de *Elaphe radiata*, una especie con un sistema de determinación sexual ZW. En cuanto a los clústers 3 y 4, identificados como LINEs y formados por lecturas de macho, su origen es desconocido, siendo una incógnita el como pueden aparecer secuencias exclusivas de machos en una especie con este sistema de determinación sexual.

6. Conclusiones

Nuestros resultados muestran que el método clásico de análisis de las familias de ADN satélite, basado en una muestra relativamente muy pequeña de sus unidades de repetición, condujo a la detección de las principales familias de estas secuencias repetitivas en el genoma, e incluso la estima de la proporción de distintas subfamilias obtenidas por ese método no está muy lejos de la producida por el análisis masivo de lecturas mediante *NGS*. De esta forma, se confirman los

resultados obtenidos con anterioridad mediante las técnicas clásicas de restricción parcial del ADN y secuenciación de productos de PCR.

El análisis mediante *NGS* y *RepeatExplorer2* es una manera fiable de estudiar las secuencias de ADN repetitivo de una especie. Este conjunto de técnicas permiten obtener una gran cantidad de información rápidamente y con un esfuerzo notablemente menor al requerido con las técnicas clásicas.

Como se ha mostrado, esta metodología no solo permite estudiar ADN repetitivo genómico, sino que permite incluso obtener genomas mitocondriales completos de especies en las que este segundo genoma no había sido secuenciado al completo (como ha sido en este caso).

7. Bibliografía

Beridze, T. (2013). *Satellite ADN*. Springer Science & Business Media.

Boissinot, S., Bourgeois, Y., Manthey, J. D., & Ruggiero, R. P. (2019). The mobilome of reptiles: Evolution, structure, and function. *Cytogenetic and genome research*, 157(1-2), 21-33.

Britten, R. J., & Kohne, D. E. (1966). Nucleotide sequence repetition in ADN. *Carnegie Inst. Wash. Year B*, 65, 78-106.

Britten, R. J., & Kohne, D. E. (1968). Repeated sequences in ADN. *Science*, 161(3841), 529-540.

Choi, I. Y., Kwon, E. C., & Kim, N. S. (2020). The C-and G-value paradox with polyploidy, repeatomes, introns, phenomes and cell economy. *Genes & genomics*, 42, 699-714.

Crepaldi, L., Policarpi, C., Coatti, A., Sherlock, W. T., Jongbloets, B. C., Down, T. A., & Riccio, A. (2013). Binding of TFIIC to sine elements controls the relocation of activity-dependent neuronal genes to transcription factories. *PLoS Genet*, 9(8), e1003699.

Crochet, P. A., Chaline, O., Surget-Groba, Y., Debain, C., & Cheylan, M. (2004). Speciation in mountains: phylogeography and phylogeny of the rock lizards genus *Iberolacerta* (Reptilia: Lacertidae). *Molecular Phylogenetics and Evolution*, 30(3), 860-866.

ELIXIR. (s.f.). *About us*. Obtenido de <https://elixir-europe.org/about-us>

ELIXIR. (s.f.). *Galaxy Community*. Obtenido de <https://elixir-europe.org/communities/galaxy>

- Feliciello, I., Pezer, Ž., Kordiš, D., Bruvo Mađarić, B., & Ugarković, Đ. (2020). Evolutionary history of alpha satellite ADN repeats dispersed within human genome euchromatin. *Genome biology and evolution*, 12(11), 2125-2138.
- Galán, P., Vila, M., Remón, N., & Naveira, H. F. (2007). Caracterización de las poblaciones de *Iberolacerta monticola* en el Noroeste ibérico mediante la combinación de datos morfológicos, ecológicos y genéticos. *Munibe (suplemento)*, 25, 34-43.
- Garcia-Porta, J., Irisarri, I., Kirchner, M., Rodríguez, A., Kirchhof, S., Brown, J. L., . . . Valero, K. C. (2019). Environmental temperatures shape thermal physiology as well as diversification and genome-wide substitution rates in lizards. *Nature communications*, 10(1), 1-12.
- Garrido-Ramos, M. A. (2017). Satellite ADN: an evolving topic. *Genes*, 8(9), 230.
- Giovannotti, M., Rojo, V., Nisi Cerioni, P., González-Tizón, A., Martínez-Lage, A., Splendiani, A., . . . Caputo Barucchi, V. (2014). Isolation and characterization of two satellite ADNs in some Iberian rock lizards (Squamata, Lacertidae). *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 322(1), 13-26.
- Hake, S., & Walbot, V. (1980). The genome of *Zea mays*, its organization and homology to related grasses. *Chromosoma*, 79(3), 251-270.
- Kapitonov, V. V., & Jurka, J. (2007). Helitrons on a roll: eukaryotic rolling-circle transposons. *TRENDS in Genetics*, 23(10), 521-529.
- Kit, S. (1961). Equilibrium sedimentation in density gradients of ADN preparations from animal tissues. *Journal of molecular biology*, 3(6), 711-IN2.
- Kubis, S., Schmidt, T., & Heslop-Harrison, J. S. (1998). Repetitive ADN elements as a major component of plant genomes. *Annals of Botany*, 82, 45-55.
- Meselson, M., Stahl, F. W., & Vinograd, J. (1957). Equilibrium sedimentation of macromolecules in density gradients. *Proceedings of the National Academy of Sciences of the United States of America*, 43(7), 581.
- Miller, O. J., & Therman, E. (2001). *Human chromosomes* (Cuarta edición ed.). Barcelona: Springer.

- Novák, P., Neumann, P., & Macas, J. (2020). Global analysis of repetitive ADN from unassembled sequence reads using RepeatExplorer2. *Nature Protocols*, 15(11), 3745-3776.
- Novak, P., Neumann, P., Pech, J., Steinhaisl, J., & Macas, J. (2013). RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics*, 29(6), 792-793.
- Pierce, B. (2002). *Genetics: a conceptual approach* (Segunda ed.). Editorial Medica Panamericana.
- Piskurek, O., Austin, C. C., & Okada, N. (2006). Sauria SINEs: novel short interspersed retroposable elements that are widespread in reptile genomes. *Journal of molecular evolution*, 62(5), 630-644.
- Remón, N., Galán, P., & Naveira, H. (2012). Chronicle of an extinction foretold: genetic properties of an extremely small population of *Iberolacerta monticola*. *Conservation Genetics*, 13(1), 131-142.
- Remón, N., Galán, P., Vila, M., Arribas, O., & Naveira, H. (2013). Causes and evolutionary consequences of population subdivision of an Iberian mountain lizard, *Iberolacerta monticola*. *PloS one*, 8(6), e66034.
- Rojo Oróns, V. (2015). Cytogenetic and molecular characterization of lacertid lizard species from the Iberian Peninsula.
- Rojo, V., Martínez-Lage, A., Giovannotti, M., González-Tizón, A. M., Cerioni, P. N., Barucchi, V. C., . . . Naveira, H. (2015). Evolutionary dynamics of two satellite ADN families in rock lizards of the genus *Iberolacerta* (Squamata, Lacertidae): different histories but common traits. *Chromosome Research*, 23(3), 441-461.
- Schildkraut, C. L., Marmur, J., & Doty, P. (1962). Determination of the base composition of deoxyribonucleic acid from its buoyant density in CsCl. *Journal of molecular biology*, 4(6), 430-443.
- Singh, L., Purdom, I. F., & Jones, K. W. (1980). Sex chromosome associated satellite ADN: evolution and conservation. *Chromosoma*, 79(2), 137-157.
- Smith, G. P. (1976). Evolution of repeated ADN sequences by unequal crossover. *Science*, 191(4227), 528-535.

- Southern, E. M. (1975). Long range periodicities in mouse satellite ADN. *Journal of molecular biology*, 94(1), 51-69.
- Thakur, J., Packiaraj, J., & Henikoff, S. (2021). Sequence, chromatin and evolution of satellite ADN. *International Journal of Molecular Sciences*, 22(9), 4309.
- Walker, P. M. (1971). “Repetitive” ADN in higher organisms. *Progress in biophysics and molecular biology*, 23, 145-190.
- Willard, H. F. (1985). Chromosome-specific organization of human alpha satellite ADN. *American journal of human genetics*, 37(3), 524.